VOLUME 106 PART B NUMBER 28 JULY 1959

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

ELECTRONIC AND COMMUNICATION ENGINEERING
(INCLUDING RADIO ENGINEERING)

SAVOY PLACE · LONDON W.C.2

Price Ten Shillings and Sixpence

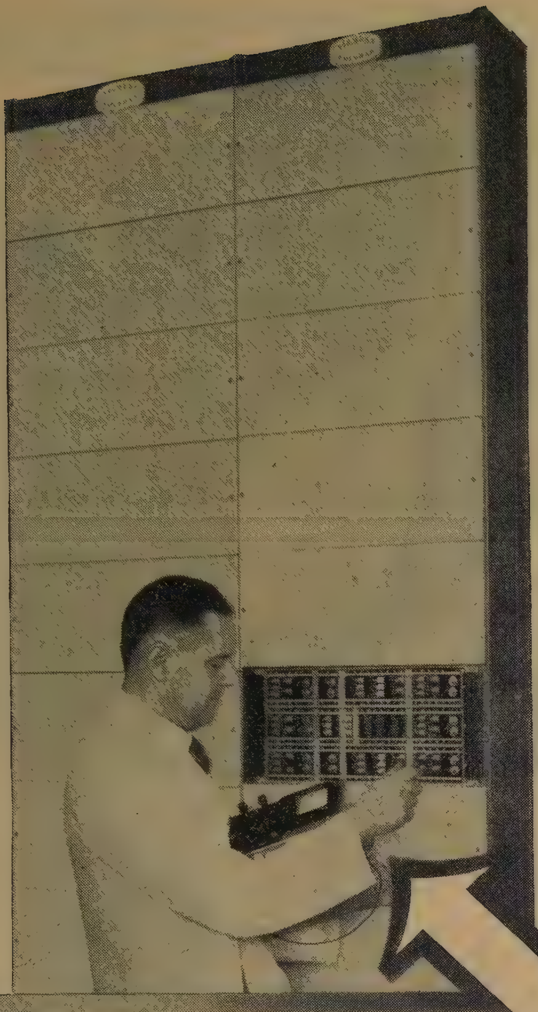
THE INSTITUTION OF ELECTRICAL ENGINEERS

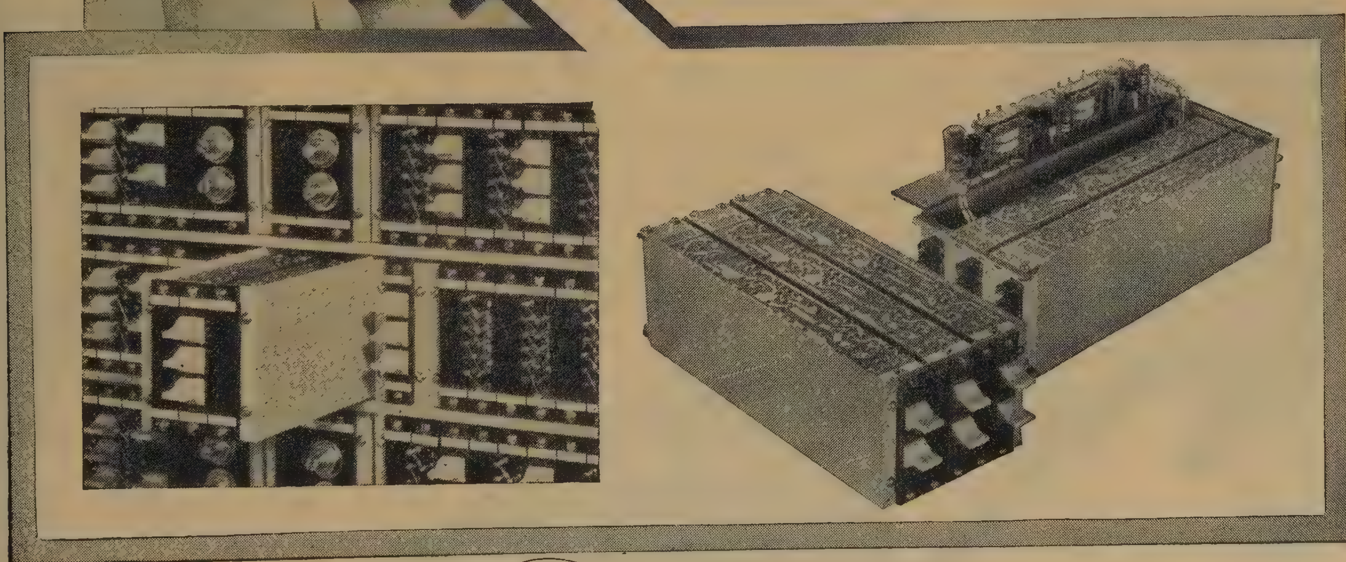
FOUNDED 1871 INCORPORATED BY ROYAL CHARTER 1921

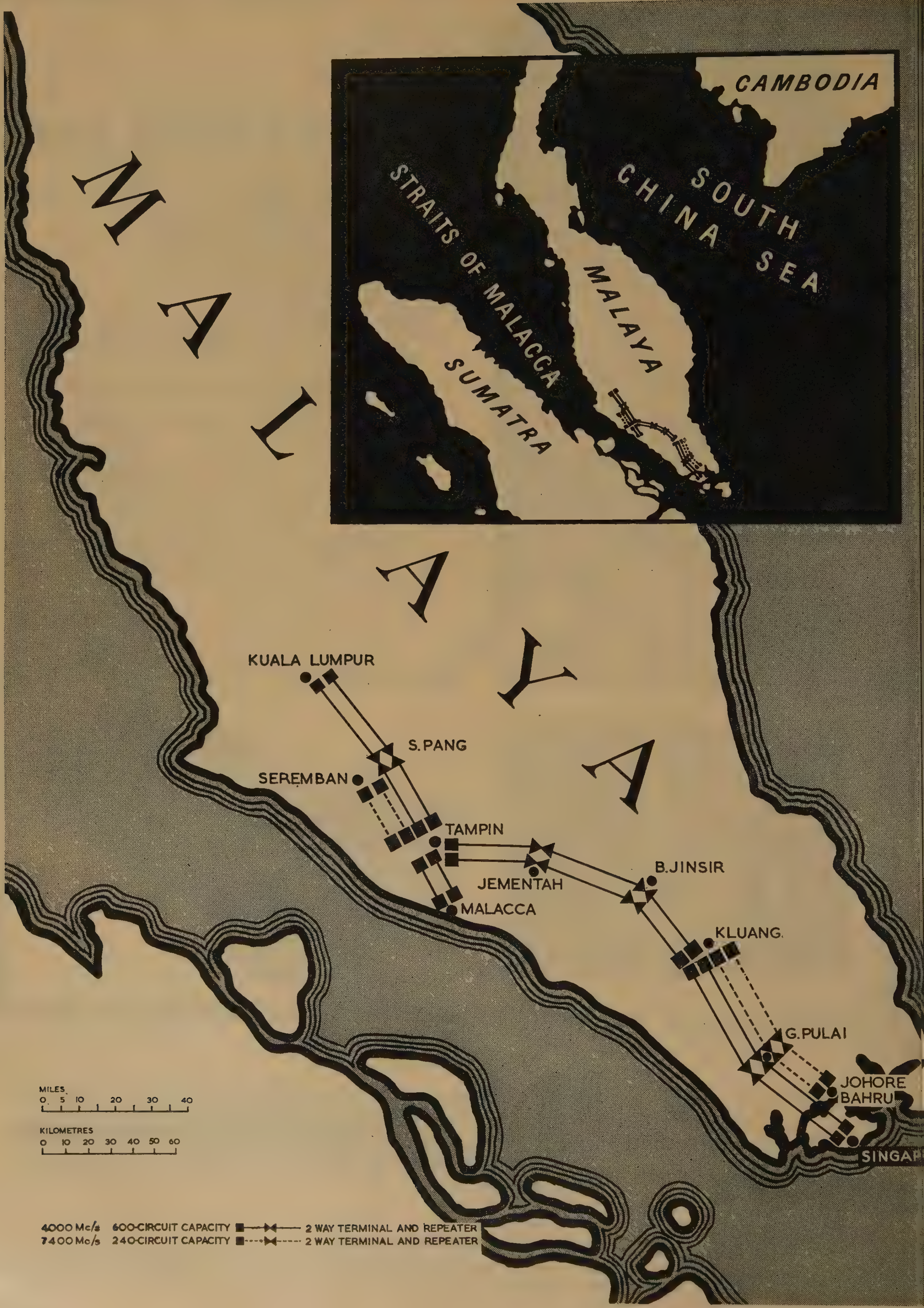
PATRON: HER MAJESTY THE QUEEN

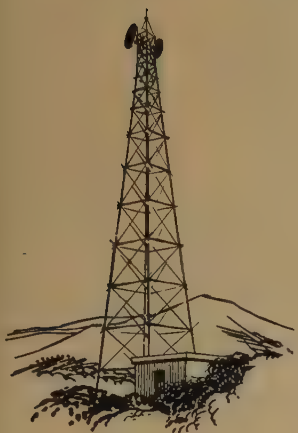

G.E.C.
Gecophone
for Overseas Administrations
G.E.C.


RADIO
CHANNELLING
EQUIPMENT R60B
• For 60 carrier channels
• Fully transistorised and compact
• C.C.I.T.T. Performance
• Advanced components, methods and design
• Inbuilt signalling equipment
This equipment combines up to 60 speech circuits, in 5 groups of twelve, into a composite signal for v.h.f. or u h.f. radio transmission. The design is new throughout and the advantages of modern components and circuitry have been fully exploited. For details request leaflet. Jel. 3302
ATE
TRANSMISSION EQUIPMENT TYPE CM.
AUTOMATIC TELEPHONE & ELECTRIC CO. LTD
STROWGER HOUSE, ARUNDEL STREET, LONDON, W.C.2.






CAMBODIA
SOUTH CHINA SEA
STRAITS OF MALACCA
MALAYA
SUMATRA
MALAYA
KUALA LUMPUR
S. PANG
SEREMBAN
TAMPIN
JEMENTAH
B. JINSIR
MALACCA
KLUANG
G. PULAI
JOHORE BAHRU
SINGAP
MILES
0 5 10 20 30 40
KILOMETRES
0 10 20 30 40 50 60
4000 Mc/s 600-CIRCUIT CAPACITY 2 WAY TERMINAL AND REPEATER
7400 Mc/s 240-CIRCUIT CAPACITY 2 WAY TERMINAL AND REPEATER



MAKING IN MILLIONS

...we have been making fuses for almost thirty years

Regd.

Most "Belling-Lee" products are covered by patents or registered designs or applications therefor.

We make millions of fuses and fuseholders in a most comprehensive range of types and sizes. Where our range—extensive as it is—does not provide for your particular needs, we are always prepared to discuss special types.

We are specialists in components. This is no empty phrase: it means that we have been engaged in the design and manufacture of precision-made components for some thirty-five years, and at Enfield we devote a very great deal of our time and space to research, forever seeking new and better methods of production, whilst ensuring that every component maintains our high standard of performance.

TERMINALS · PLUGS & SOCKETS
FUSES · INTERFERENCE FILTERS
THERMAL DEVICES · RECEIVING AERIALS

BELLING & LEE LTD

GREAT CAMBRIDGE ROAD, ENFIELD, MIDDX., ENGLAND

Telephone: Enfield 3322 · Telegrams: Radiobel, Enfiel

TOROIDAL SUPPRESSION CHOKES

used in all the major airliners including the COMET, VISCOUNT and BRITANNIA

With the greater use of radio and navigational aids on ships and aircraft, it is increasingly important to suppress interference likely to arise from other electrical equipment which is also an inseparable part of modern transport. S.T.C. with over 40 years experience in the development of high grade magnetic materials, now introduce a selection of the many types of chokes which are manufactured for industrial, marine and aeronautical applications.

Actual Size C.1.

RANGE OF SIZES AND WEIGHTS

The following table lists the size, weight and normal range of inductance and current rating of several S.T.C. Toroidal Chokes:

SIZE CODE	SIZE Outer Diam: cm	Inner Diam: cm	Length cm	Weight gm	INDUCTANCE & CURRENT RATING (μH and Amp. AC or DC)
C1	1.3	0.7	0.52	3.5	from 10 μH(½ amp) to 1000 μH(0.1 amp)
C2	2.3	1.2	0.80	14.5	,, 100 ,, (1½ ,,) to 1000 ,, (½ ,,)
C3	2.95	1.4	1.3	50.0	,, 100 ,, (2½ ,,) to 1000 ,, (½ ,,)
C4	4.1	2.4	1.3	85	,, 100 ,, (4 ,,) to 2000 ,, (1½ ,,)
C5	4.8	2.7	1.4	130	,, 100 ,, (5 ,,) to 2000 ,, (2 ,,)

COMPONENTS GROUP

Standard Telephones and Cables Limited

Registered Office: Connaught House, Aldwych, London, W.C.2

MAGNETIC MATERIALS DEPT: NORTH WOOLWICH · LONDON · E.16

vast experience in the science of telecommunications

As a result of years of experience in the world of telecommunications, we have an unsurpassed wealth of knowledge of this science, which is freely offered to you as part of the incomparable Siemens Ediswan service.

- Comprehensive Telecommunication Systems
- Public Telephone Exchanges
- Private Telephone Systems
- Telephone Instruments and Apparatus
- Carrier Telephone Systems
- F.M.V.F. Telegraph Equipment
- V.H.F. Radio Telephone Terminals
- Marine Signalling Apparatus
- Marine Radio Equipment
- Control & Indicating Apparatus

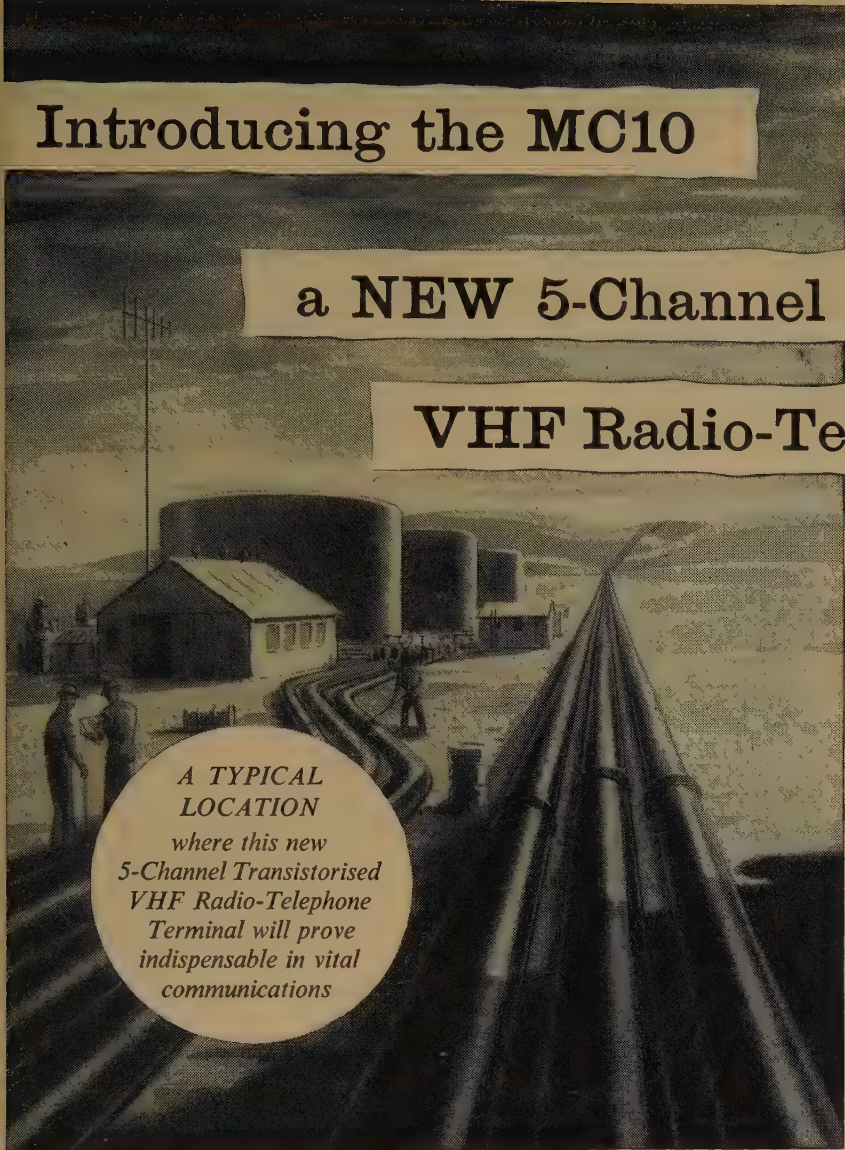
SIEMENS EDISON SWAN LTD.
An A.E.I. Company
Telecommunications Division, Woolwich, London S.E.18
Telephone: Woolwich 2020

TA2306

Introducing the MC10

a NEW 5-Channel Transistorised

VHF Radio-Telephone Terminal



Compact design which at low cost provides five high-grade telephone circuits

Radio

Frequency range 156 - 184 Mc/s
(other frequency ranges are available)
Transmitter power output 30 watts
Deviation 75 kc/s
Receiver Noise Factor 8db
All characteristics of the transmitter and receiver conform to CCIR specifications

Carrier Telephone Equipment

5 Telephone channels · 4 kc/s spaced
Equipped with Out of Band Signalling
Facilities for dialling, Ringdown or junction working
Printed Wiring · Plug-in Units
Crystal frequency control
Resin cast components

The result of co-operative enterprise between two great organizations

Redifon

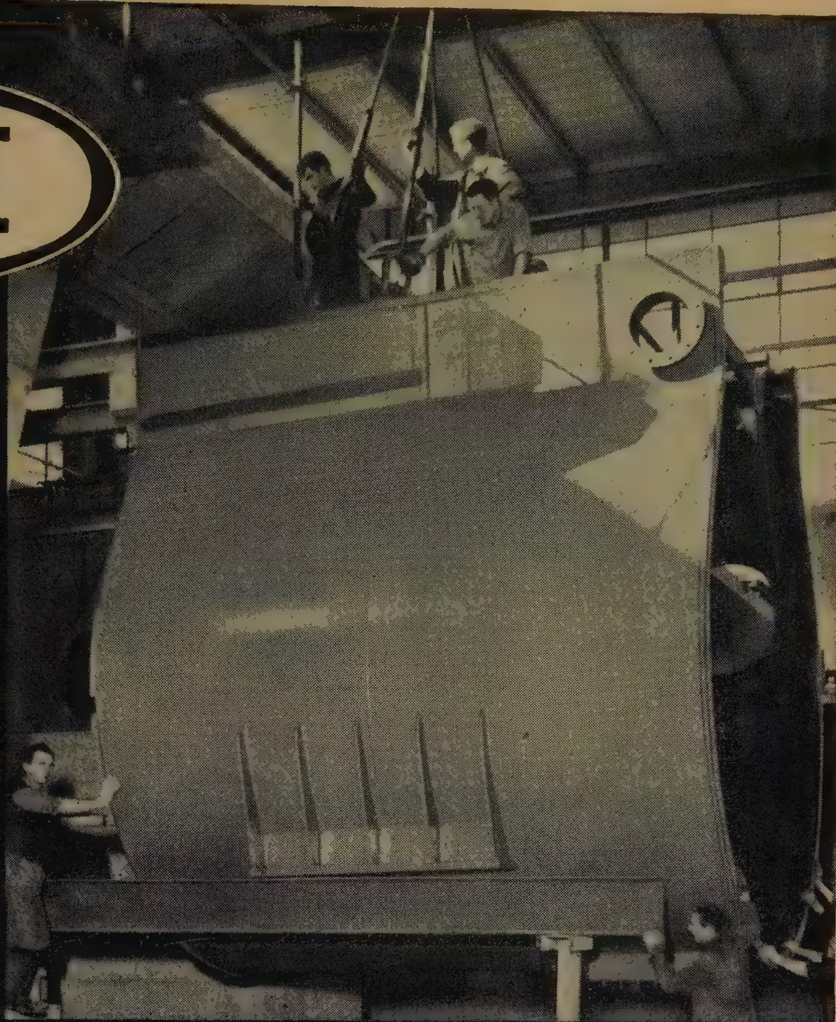
SIEMENS EDISWAN

REDIFON LIMITED
Communications Division, Wandsworth, London, S.W.18
Telephone : VANdyke 7281
A Manufacturing Company in the Rediffusion Group

SIEMENS EDISON SWAN LIMITED
Transmission Division, Woolwich, London, S.E.18
Telephone : Woolwich 2020
An A.E.I. Company

AEI

COMPLETE TURBINE-HOUSE EQUIPMENT



One unit of a three-shell surface condenser being placed in position at Berkeley, the commercial nuclear power station now under construction in Gloucestershire by the A.E.I.—John Thompson Nuclear Energy Company.

FOR BERKELEY NUCLEAR POWER STATION

A general view of three condenser shells in position at Berkeley.

The entire turbine-house heat exchange equipment for the new power station at Berkeley is being supplied by the A.E.I. Turbine-Generator Division, and manufactured at their factory at Germiston Glasgow.

Four three-shell main condensers — one for each of the four 83 MW mixed-pressure steam turbine-generators also being supplied by A.E.I.

Two-stage feedwater heating equipment for each turbine-generator.

Two 'dumped' steam condensers to handle surplus steam generated when commissioning the reactors, or in the event of a sudden shut down.

Other equipment in manufacture includes the steam jet air ejectors, water/water heat exchangers for the generator cooling circuit and general gas circulator service, oil coolers and generator hydrogen cooler units.

ASSOCIATED ELECTRICAL INDUSTRIES LTD.

TURBINE-GENERATOR DIVISION

incorporating the interests of M-V and B.T-H. Manchester and Rugby.

WORKS AT MANCHESTER AND RUGBY, ENGLAND. GERMISTON, SCOTLAND. LARNE, NORTHERN IRELAND.

B/A 804

INDEX OF ADVERTISERS

THE PROCEEDINGS OF
THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 106. PART B. No. 28. JULY 1959

621.395.342

The Institution of Electrical Engineers
Paper No. 2795 R
Nov. 1958
©

SUBSCRIBER TRUNK DIALLING

The Scheme for Full Automation of the Telephone Service in the United Kingdom

By D. A. BARRON, M.Sc., Member.

(*The paper was first received* 8*th September, and in revised form* 21*st October*, 1958. *It was published in November*, 1958, *and was read before* THE INSTITUTION 22*nd January, the* SOUTH MIDLAND RADIO AND MEASUREMENT GROUP 26*th January, the* NORTH-EASTERN RADIO AND MEASUREMENT GROUP 2*nd February, the* NORTH-WESTERN CENTRE 3*rd March, the* EAST MIDLAND CENTRE 7*th April, the* SHEFFIELD SUB-CENTRE 15*th April, the* NORTH STAFFORDSHIRE SUB-CENTRE 16*th April, and the* SOUTH-EAST SCOTLAND SUB-CENTRE 29*th April*, 1959.)

SUMMARY

Plans have been prepared by the Post Office to enable subscribers to dial their own trunk calls. This is called 'subscriber trunk dialling' (s.t.d.).

The new service will be opened at Bristol Central Exchange in December, 1958, and will be extended as rapidly as resources permit. By the end of 1961, subscriber trunk dialling is expected to be available in over 100 centres.

A national numbering scheme has been prepared, and ultimately a long-distance call from any part of the United Kingdom will be obtained by dialling the national number of the required subscriber. Local numbers will be unaffected.

Exchanges have been divided into groups for charging and routing purposes. The charging and routing will be controlled by register-translator equipment (Grace), an electronic form of which will be used for the larger installations.

In s.t.d. areas, new methods of charging for calls will be introduced. All calls will be timed, and the unit fee will be reduced to 2d. The existing 3 min minimum charge will not apply to s.t.d. calls. Instead, unit fees will be recorded, each unit paying for a period of time which depends on the distance of the call. Both trunk and local units will be registered on existing subscribers' meters, and a bulked bill will be rendered. Private meters will be rented to subscribers who require them. Accounts will be sent to s.t.d. subscribers quarterly, and the preparation of accounts will be increasingly mechanized. A new coin box has been developed so that trunk calls in s.t.d. areas can also be dialled by coin-box users.

(1) INTRODUCTION

(1.1) The Extent to which Dialling by Subscribers has Hitherto been Possible in the United Kingdom

In the United Kingdom, as elsewhere, the earliest telephone exchanges were manually operated, but since the opening of the first automatic exchange in 1912, there has been an ever-increasing proportion of local automatic working.

Inter-dialling by subscribers between automatic exchanges has been progressively introduced, but has been restricted to 'local' calls, i.e. the shorter-distance calls which have been untimed and charged for in single units or multiples of that unit. By 1936, with the advent of multi-metering, the local dialling range had been extended to a limit of 15 miles chargeable distance, the charges being 1 unit for 0–5 miles, 2 units for 5–7½ miles, 3 units for 7½–12½ miles, and 4 units for 12½–15 miles. The latest change, in January, 1958, resulted from the combination of exchanges for charging purposes into groups, the unit-fee local-call area being extended to cover the home and adjacent groups. The permissible local dialling range for a single unit was thereby increased to an average of 17 miles, but it often extends to some 25 miles or more, depending on the configuration of particular groups.

(1.2) Normal Stages in the Mechanization of Trunk Operation. Definition of Subscriber Trunk Dialling

Hitherto, the longer-distance calls, which have been charged for according to the time taken as well as the distance, have always had to be set up by an operator. It is with these calls that the paper is primarily concerned. They will be referred to as trunk calls, a term well established and understood in this country, and broadly equivalent to a toll call in the United States, or a *conversation interurbaine* in France. Some indication of the size of the problem is given by the fact that some 321 million inland trunk calls were made in the United Kingdom in 1957.

The method of handling trunk calls logically evolves through three main stages, which may be co-existent in different parts of a system as the necessary changes are gradually introduced. These three stages are shown in simplified form in Fig. 1. In the interests of clarity, no intermediate switching points have been shown.

In Stage 1, operators are used at all trunk switching points. In Stage 2, trunk calls are completed by the controlling operator without the assistance of any other trunk operator. In Stage 3, trunk calls can be dialled by a subscriber without requiring the

Mr. Barron is in the Post Office Engineering Department.

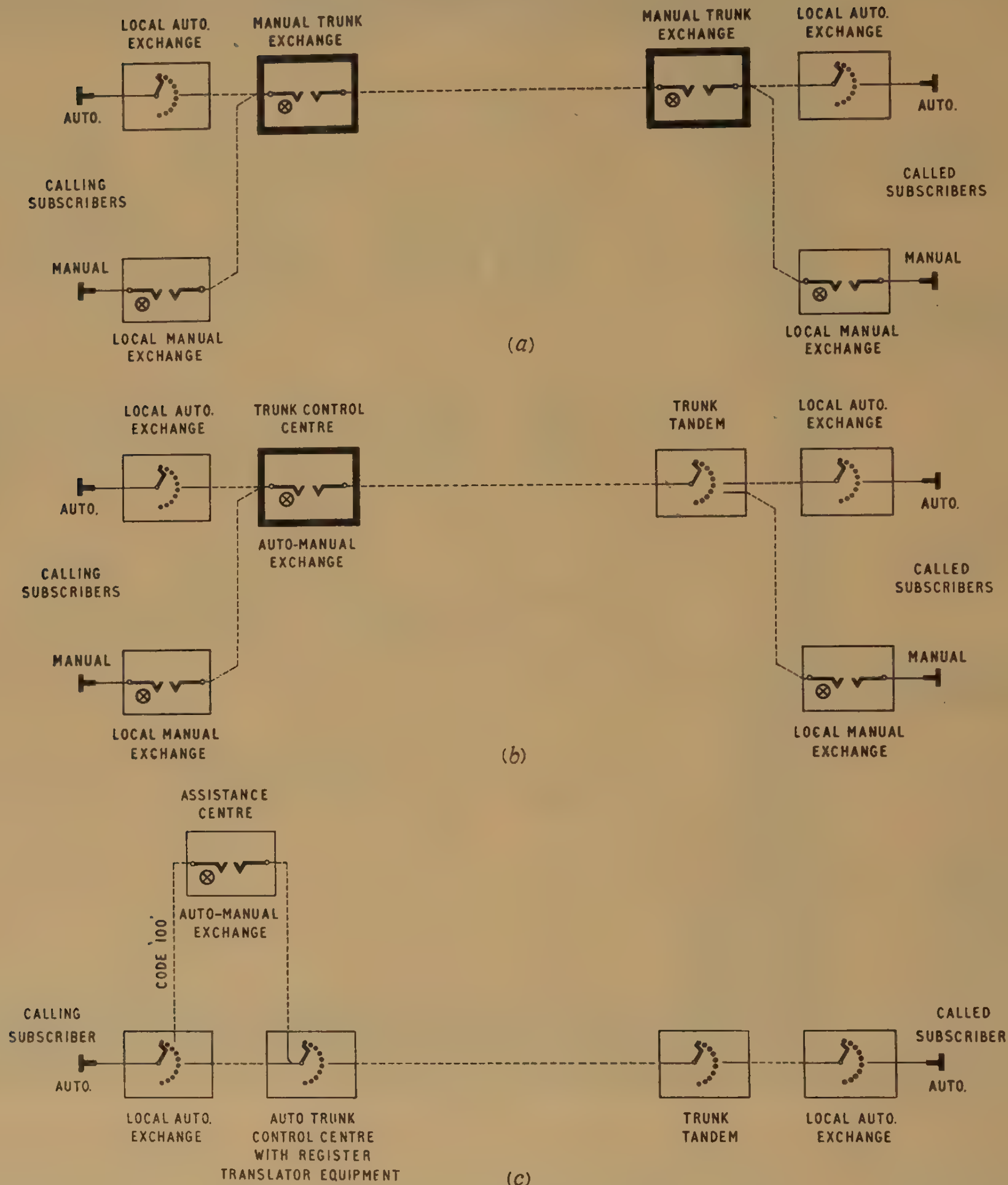


Fig. 1.—Logical stages in the development of trunk working.

(*a*) Stage 1. Manual operation.
(*b*) Stage 2. Semi-automatic operation.
(*c*) Stage 3. Fully automatic operation (subscriber trunk dialling).

intervention of an operator. This has come to be known in the United Kingdom as *subscriber trunk dialling* (and in America as *direct distance dialling*).

(1.3) The Stage Reached in Other Countries

Progress varies within countries and, for international traffic, between countries. Most countries well developed telephonically have introduced, or are planning for, subscriber trunk dialling in a form suited to their requirements. It is not possible to attempt a comprehensive survey, but a few examples are given.

Belgium.—Some 81% of the total (about 1 million) telephones are connected to automatic exchanges, and about 68% of the long-distance calls are completed either automatically or semi-automatically.

Federal Germany.—About 98% of the total (about 4·8 million) telephones are now automatic, and it is expected that, by the end of 1960, the remaining manual exchanges will have been converted. By April, 1958, the proportion of long-distance calls dialled by subscribers was 67·7%.

France.—About 55% of the total (about 3·7 million) telephones are now connected to fully automatic exchanges. In addition, about 17½% are connected to rural semi-automatic exchanges. A limited measure of fully automatic telephone operation between groups (*service interurbain automatique*) was introduced as long ago as 1939, and automatic service was also provided on certain long-distance routes, chiefly outgoing from Paris, in 1951. A national numbering plan was established in 1955. Some 25% of the total traffic between groups (*trafic interurbain*) is now completed on a fully automatic basis.

Netherlands.—Some 96% of the total (about 1·5 million) telephones are now connected to automatic exchanges, and most of them have facilities for nation-wide trunk dialling. About 92% of long-distance traffic was being completed automatically by 31st December, 1957.

Sweden.—Some 82% of the total (about 2·5 million) telephones are connected to automatic exchanges. About 50% of the long-distance calls are now dialled by subscribers.

Switzerland.—The mechanization of local and trunk telephone service has been substantially complete for many years, and tribute should be paid to the valuable pioneer work carried out by the Swiss P.T.T. Department. There are about 1·4 million telephones.

United States and Canada.—About 90% of the total (about 65 million) telephones are connected to automatic exchanges. In 1957 nearly 2750 million toll calls were handled, of which about 55% were dialled by operators and 20% by subscribers. Direct distance dialling is now available to over 5 million subscribers, and it is anticipated that the scheme will be substantially completed within 10 years.

(1.4) United Kingdom Position and Programme

The number of telephones in the United Kingdom is about 7·5 million. Some 80% are served by automatic exchanges. The largest cities are served by linked numbering schemes, and the biggest of these—the London Director Area—alone provides fully automatic interconnection facilities between some 1·8 million telephones, i.e. more than the national total in smaller countries such as Switzerland and the Netherlands.

Planning to permit operator dialling of trunk calls commenced as long ago as 1934, and the system was introduced during 1939–40 in spite of the war. Designed basically for single link dialling via circuits in the operators' outgoing multiple, its application to routes between the large centres resulted in substantial economies.

From 1944 a limited amount of tandem operation was introduced, and during 1944–54 there was a gradual extension of semi-automatic working, so that, by 1954, about 33% of all trunk traffic was handled on this basis. The development of a more comprehensive scheme, providing for automatic selection of the outgoing circuits and for multi-link tandem operation, then permitted further extension. This scheme, known as *trunk mechanization*, was operated on a straightforward basis, the codes dialled by operators for any particular objective exchange varying according to the point of origin. By mid-1958 trunk units at London, Swansea, Birmingham, Manchester, Chester and Carlisle were in operation, and some 50% of the full-rate trunk traffic was handled on a semi-automatic basis.

It is, however, only by the extension of fully automatic operation to the trunk service that the best and fastest service can be provided at the lowest possible cost, and subscriber trunk dialling has been enthusiastically welcomed by the great majority of subscribers in all countries in which it has been introduced.

Planning for subscriber trunk dialling in the United Kingdom has involved the establishment of a national numbering plan, and the development of automatic equipment to replace the operator's functions, notably for the control of routing and charging. Radical changes in the tariff structure became practicable, and the savings in the costs of operating the system enable reductions to be made in s.t.d. areas in the charges both for local and subscriber-dialled trunk calls. These improvements should do much to stimulate increased use of the trunk service, and could not have been contemplated unless an adequate number of trunk circuits could be made available at reasonable cost. That this position has been reached is due to the considerable advances made in the line transmission field in the post-war years. The scheme will be opened at Bristol Central Exchange in December, 1958, and it is hoped to extend it to over 100 other centres by the end of 1961. By 1970 it is expected that not less than 75% of all trunk calls will be dialled by subscribers.

(2) DESIGN PRINCIPLES

(2.1) General

Subscriber-trunk-dialling schemes vary considerably in different countries according to the size of the territory, the nature and layout of switching and line plant, and the numbering, routing and tariff backgrounds. For example, the networks in some countries have been developed in such a way that the territory is divided into a number of groups of exchanges, each group being served by a common range of local numbers. Another arrangement is to provide a form of decade numbering, which, in principle, permits any objective exchange to be obtained on a straightforward dialling basis, but loses routing flexibility. In practice, some form of discriminating arrangement is usually provided in the latter systems to avoid unnecessarily circuitous routings.

Radical differences also exist in the methods adopted for recording information about trunk calls for billing purposes. In some countries, notably the United States and Canada, the s.t.d. arrangements provide for the automatic recording of details of individual trunk calls so that they can be itemized on subscribers' accounts. Factors leading to this requirement were the very large distances involved and the absence of subscribers' meters in many areas where a flat-rate local tariff is in force. In most European countries, however, some form of common metering of both local and trunk calls, with a bulked account, is preferred. The tariffs, and the form and frequency of accounting, vary considerably. There are also, as would be expected, considerable variations in plant design, layout and function. The Comité Consultatif International Télégraphique et Téléphonique has issued certain information and recommendations, directed primarily to easing the interworking problem with international subscriber dialling, which refer chiefly to such matters as the following:

(*a*) The number of digits in the national number.
(*b*) International country code lists.
(*c*) The desirability of arranging that the first one or two digits of a national number should have a geographical significance, in cases where a country is divided into more than one zone for international charging purposes or has more than one point of entry for international calls. (These considerations do not apply to the United Kingdom.)
(*d*) The standardization, and preferred method of indication, of prefix digits.
(*e*) The use of letters in national numbers.

Subject only to meeting these general recommendations as far as possible, each country develops the particular scheme best suited to its national requirements.

(2.2) The Principles Underlying the United Kingdom Scheme

The design of the s.t.d. scheme for the United Kingdom has been based on the following principles:

(*a*) It should be simple and economical to provide and maintain.
(*b*) It should be easy for subscribers to understand and use.
(*c*) It should be capable of being progressively introduced without requiring extensive changes to existing plant.
(*d*) It should provide features appropriate to the new method of working rather than attempt to simulate outmoded aspects of the existing manual system.
(*e*) It should provide sufficient flexibility to meet the routing requirements of the existing network and not be restrictive on future line-plant planning.
(*f*) It should have adequate capacity to meet the anticipated ultimate requirements.

The application of these principles, in the light of the conditions existing in the United Kingdom system, has resulted in a scheme which differs in a number of respects from that of any other country. In the following Sections, the numbering, charging and routing problems and their technical solutions are discussed in detail.

(3) NATIONAL NUMBERING

(3.1) General

The essential requirement, in each country, is to provide each subscriber with a unique national number. When mechanization is complete, long-distance calls from any part of a country to any particular subscriber in that country can be obtained by dialling his national number. The national number, preceded by some digits to indicate the required objective country, can also be used for fully automatic international calls.

Each subscriber already has a local telephone number identifying his line, or group of lines, in the local exchange system. A national number can therefore be built up by using additional digits, preceding the local numbers, to identify the individual exchange systems.

The local number may be in (*a*) a local system in which a number of exchanges are linked together in one common numbering scheme, or (*b*) a separate exchange with independent numbering. Linked-numbering areas vary considerably in size in different localities and in different countries. In the United Kingdom, local linked numbering schemes are used only in compact multi-exchange areas where there is a large volume of inter-exchange traffic. Examples are the director systems in the largest cities (London, Birmingham, Edinburgh, Glasgow, Liverpool and Manchester) and the non-director systems in other large provincial cities (e.g. Bristol, Leeds). In some countries, where the whole territory has been divided into linked numbering areas, such areas may be very large. For example, about 100 'numbering plan' areas cover the whole of the United States and Canada. On the other hand, Switzerland has been divided into about 50 such areas. It is the general practice that, in linked numbering schemes, the first few digits identify the particular exchange, and the other digits the particular subscriber on that exchange. Systems like those in London (MAYfair 1234) and New York (MUrray Hill 6–1234) are essentially 7-digit linked numbering schemes. The first three digits identify the exchange, some letters being used instead of figures simply to help the caller to remember the number and dial it correctly. Again, in Bristol, which is a mixed 5- and 6-digit area, all numbers on the central exchange begin with the figures 2 or 9.

In a national numbering plan, a common code can be used to identify the group of exchanges forming a linked numbering scheme. In countries whose territory is completely served by such linked numbering plan areas, therefore, the creation of a national numbering plan consists essentially of the allocation of suitable codes to identify the separate areas. The code and the local number together comprise a subscriber's national number. For example, to call a subscriber in Geneva from Zürich one might dial 022.34.56.78, the code for the Geneva group being 022, and 34.56.78 the local number in that group.

(3.2) The United Kingdom National Numbering Plan

It would have been undesirable and costly to attempt to rearrange the United Kingdom network so that the whole country was covered by linked numbering plan areas as a prerequisite to national subscriber dialling. Not only would considerable plant rearrangements and additions have been necessary, but many local numbers would have had to be changed, and more digits dialled for a large number of local calls. The preferred arrangement was to leave local numbers undisturbed, the question of introducing further local linked numbering schemes being reserved for future decision in the light of plant economics and other considerations. It was therefore necessary to design a national numbering plan which would permit the identification of individual exchanges, as well as linked numbering areas, any extra digits necessary being in the national code and not in the local number. Such a plan has been evolved.

Local telephone numbers in the United Kingdom vary, on standard automatic exchanges, from three to seven digits.

A short national code is desirable where the local number is already long, especially as the largest volume of trunk traffic goes to such areas. A longer national code may then be needed for routing calls to exchanges with short local numbers. Although, in this way, national numbers tend towards a uniform length, there is little service advantage in uniformity, and the cost of arranging for the routing and charging equipment to handle a varying number of digits is not high. Moreover, a much greater saving in equipment costs can be achieved by keeping local numbers short and allowing for more digits to be added as the local system grows.

The United Kingdom scheme has therefore been designed so that, for routing purposes, the exchanges are arranged in groups. Each group has its own group code which, in most cases, is composed of a prefix digit (which the subscriber must dial to get access to the s.t.d. equipment) and three other digits. National numbers will have eight, nine or ten digits including the prefix. The only suitable single-figure code which could be made available for use as a prefix digit at all exchanges was the digit '0'. The code has traditionally been used for calls to the operator, and it is appropriate that it should be used in the new plan to obtain access to the equipment which has been designed to replace the operators' functions. With increasing mechanization, the number of calls to operators will decrease, and the new code for this purpose will be '100'.

Fig. 2 illustrates the general principle upon which exchange

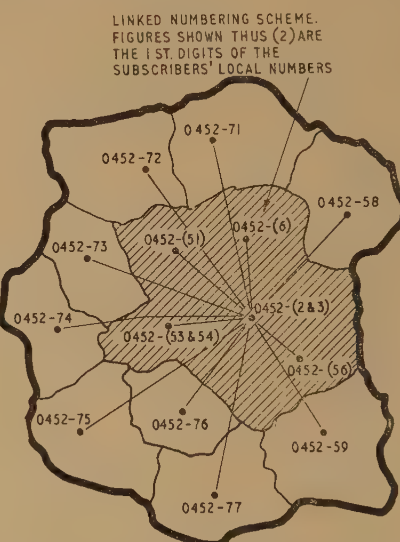


Fig. 2.—General principle used for numbering exchanges in groups.
Group code 0452

codes can be allocated so that certain digits are common to all exchanges in the group and thus constitute the group code. In this example, the central part of the group is served by a linked numbering area. The code 0452 identifies that area, and the individual exchanges in the linked numbering area are identified by the early digits of the local number in the manner already described. Other exchanges in the group, outside the linked numbering area, are separate exchanges with independent numbering. They are identified by digits added to the group code, these digits being chosen so that they do not clash with early digits of the local numbers in the linked numbering area. The

codes shown are all numerical, but the same principle could apply if a combination of letters and figures were used.

(3.2.1) The Use of Letters in National Numbers.

The C.C.I.T.T. recommends that, for the international fully automatic service, national numbering plans should preferably not involve the use of letters. This recommendation was made to avoid difficulties in countries where letters are not normally used. It has been recognized, however, that in countries where letters are already in use for local numbers their use for national purposes may be unavoidable, and it may even be desirable for national reasons to extend such use. Arrangements can be made, e.g. by using letter/figure cross reference, to enable numbers containing letters to be dialled from other countries not using lettered dials.

Letters are already used in the local numbers of the director areas London, Birmingham, Edinburgh, Glasgow, Liverpool and Manchester, and these areas include more than one-third of all the telephones in the United Kingdom. For this reason, United Kingdom subscribers given s.t.d. facilities in other areas must be supplied with dials having letters. Tests have shown that there is little to choose between random letters and figures in dialling time or accuracy, but experience has amply demonstrated the value of significant letters—the first three of the exchange name in London, etc.—in helping subscribers to dial correctly, to avoid transposition of digits, and to remember numbers. It has therefore been decided to use letters generally in British national numbers. In the group codes for the non-director areas (the general case shown in Fig. 2), two of the figures will be replaced by letters, so chosen that they are usually the first two letters of the name of the major town, or, where this is not possible, of some suitable district name.

Because of their size and importance the director area groups will be given shorter codes, and as the local numbers already include letters, the group codes will be wholly numerical. In these areas the codes will be obtained by adding to the numerical group codes the three letters which, at present, identify each individual exchange.

As a result of the application of these principles, the numbering plan will be of the form shown in Table 1.

available for use. However, there are a number of practical considerations which limit the use that will be made of this capacity.

The division of the territory into groups, the non-director groups being identified by the first three effective digits, means that the theoretical number capacity is equally divided between them (each group absorbing one million numbers), and that spare capacity in one group cannot be used in another. But many groups in sparsely telephoned territory contain very few lines—the Foula group contains only one—and these can never use much of their theoretical capacity.

In most parts of the country, including quite large towns, national numbers with eight effective digits will last as long as can be foreseen. Here, again, much of the theoretical capacity will not be used.

The linked numbering schemes in certain large provincial cities form the largest non-director groups. In these cases, national numbers can have nine effective digits (i.e. three in the code and six in the local number), but practical limitations, such as the inability to use local numbers beginning with the digits '0' and '1', reduce the effective number capacity to about 600 000. Although eventually, when there may be as many as 20 million exchange connections in the country, the largest of these groups (Leeds has 48 000 lines at present) may have grown to 200 000 lines, there will still be an ample margin of numbering capacity. This would not have been provided by a scheme with only eight effective digits.

Each director area may be regarded as a 7-digit linked numbering scheme with a theoretical capacity of 10 million numbers. The practical capacity is restricted to somewhat less than half this figure by the inclusion in local numbers of codes derived from the first three letters of exchange names. Even so, the capacity is more than adequate to meet the forecast development over the next 100 years.

Most of the available numbering capacity is distributed between the groups, and comparatively few group codes will remain unallotted. As from 1st January, 1958, the British network has been divided into 639 charging groups, but, for numbering and routing purposes, some of these will be divided, each part being given its own group code. A group or part of a group with its own code is known as a 'number group', and there will be about 700 of these altogether. Each number group will have a 4-digit code of the form 0 AB2, of which there could be

Table 1

TYPICAL NATIONAL NUMBERS

Type of exchange	Typical number at present	Typical National number
London director	ABB 1234	01 ABB 1234
Provincial director	MID 1234	021 MID 1234
Non-director switching centre or satellite in associated linked numbering scheme		
4-digit	Taunton 2345	OTA3 2345
5-digit	Leeds 22345	OLE2 22345
6-digit	Adel (Leeds) 672345	OLE2 672345
Other non-director exchanges (including unit automatic exchanges)		
3-digit	Henlade 256 (near Taunton)	OTA3 64 256
4-digit	Shoreham by Sea 2345 (near Brighton)	OBR3 7 2345
4-digit	Droitwich 3223 (near Worcester)	OWO5 73 3223
5-digit	Whitley Bay 22516 (near Newcastle)	ONE2 6 22516

(3.2.2) The Capacity of the Selected Scheme.

A numbering scheme with nine effective digits could have a theoretical capacity of 10^9 numbers if all digits were freely

$9 \times 9 \times 10 = 810$. There will, exceptionally, be the shorter codes; 01 for the London group, and nine codes of the form 021, of which five will be allotted initially to provincial director areas. Thus the capacity for group codes of the forms described is $810 + 1 + 9 = 820$. The margin is regarded as sufficient, because the further subdivision of number groups should not be necessary. Indeed rather the reverse is expected, since the growth of the local system will tend to justify the extension of local linked numbering schemes to absorb separate number groups which have been established in the same charging group.

A section of territory, divided into typical groups, is shown in Fig. 3.

Fig. 3.—Section of the country showing layout of charging groups and their subdivision where necessary into numbering groups.

—— Charging group boundary.
····· Numbering group boundary.

The section shown is the local fee area for the shaded charging group. The codes are typical ones only.

(3.2.3) Local Code Dialling.

Where linked numbering schemes are not in operation, access to other exchanges in the same group will be obtained by dialling local inter-exchange codes, usually of two digits. The volume of traffic routed in this way may be large, and the use of local codes rather than the longer national codes will be more economical. Where transmission and signalling conditions permit, local codes may also be used for access to exchanges in adjacent groups, to which the call charge, under group charging, is the same as for calls within the group.

(3.2.4) Publishing the Codes.

The new national codes, together with local codes where they are used, will be issued to subcribers in booklet form. The booklet issued for a particular group will be reissued from time to time as the range of dialling from that group is extended by the opening of new exchanges and trunk switching centres throughout the country.

(4) AUTOMATIC CONTROL OF ROUTING AND CHARGING

(4.1) The Principles used in the Register-Translator Control

In order (*a*) to be able to obtain access to a distant subscriber by dialling always the same digits—i.e. his unique national number—irrespective of the point of origin of the call, and (*b*) to obtain routing flexibility, it is necessary to divorce numbering from routing, and to interpose a device which translates the digits dialled into the most suitable form for routing the call. Such devices are well known, and are called register-translators. The register receives and registers the numbers dialled by subscribers, and the translator provides routing information by translating those digits of the dialled number which indicate destination into routing information. The British Post Office director is a familiar example. With register-translators there are two extremes in the method of routing control. On the one hand, a register-translator may provide routing information just sufficient to locate the appropriate group of trunk or junction circuits outgoing from the exchange in which it is located. On the other hand, it may provide enough information to enable every call to be set up right through to its destination, regardless of the number of intermediate switching centres that may need to be traversed. The former type, known as the 'own-exchange-only' (o.e.o.) register-translator, needs only to provide sufficient routing information to cover the possible outgoing routes. This implies that, for calls not completed over direct links between the calling and called exchanges, part or all of the national code must be repeated forward to other registers located at intermediate switching centres. With the alternative type, the 'right-through' register-translator (the British director is of this type), no intermediate registers are required but the translators must contain complete routing information. It can be shown that, for a national scheme covering the United Kingdom, the total equipment costs are less for a layout based on o.e.o.-type translation than they would be if 'right-through' control were employed. Another important feature of the o.e.o. method of routing which has influenced its selection is the ease with which localized routing changes can be made without repercussions throughout the system.

In practice, the o.e.o. principle need not be applied too rigidly where local circumstances permit economic variations. For example, a register-translator can readily be designed to complete a limited number of 'right-through' routings to terminal exchanges via conveniently placed intermediate exchanges without requiring much additional equipment in the translator.

The early s.t.d. installations will use only controlling register-translators, which will give subscriber-dialling facilities over all the direct routes and some indirect routings. Later, subscriber dialling over indirect routings will be extended by providing, at suitable switching points, o.e.o.-type transit and terminal registers. Although the register-translator is primarily a routing control device, it will also be arranged to give charging information, determined from the code allotted to the charging group within which the wanted exchange is located. In general, both the charge rate and the route can be decided by examination of the first one, two or three digits received, but limited provision is made for examining up to the fifth digit to permit economies in routing by the use of direct routes where appropriate.

Typically, the functions of a controlling register-translator are:

(*a*) To determine the route and charge for any call.
(*b*) To use the routing information, together with such digits of the national number as are necessary, to operate switches in the originating and distant exchanges in order to set up connection either to the required subscriber or to a transit or terminal register as necessary.
(*c*) To transfer the charging information to a metering-control relay set associated for the duration of the call with the calling line.

The equipment is known as 'Grace' (group routing and charging equipment).

(4.2) Design Techniques Employed

Advantage has been taken in the design of the controlling register-translator to use electronic switching techniques for some of the larger installations. Comparative evaluation of these techniques on the basis of cost and reliability will be obtained, because electro-mechanical register-translators giving the same facilities have also been developed and will be used in appropriate cases.

The electronic type of equipment developed for use in non-director exchanges uses cold-cathode tubes interconnected by selenium-rectifier gates for storage, counting and code identification. Each digit store comprises a group of five tubes on which the digit is stored in '2 out of 5' code. The translator is common to 40 registers, to which it is allotted in turn via an electronic connector for 16·6 millisec every 667 millisec. During its allotted period a register requiring translation information marks into the translator its stored digits and an indication of the stage reached in its sending programme, and receives from the translator the appropriate translation digit. To minimize storage requirements in the register, one digit only is supplied by the translator each time, the register accepting the next digit during the inter-train pause following pulsing-out of the previous digit.

the register-translator designed for the director application, use has been made of the magnetic-drum technique. A 9 in nickel-coated aluminium drum forms the system memory, and individual registers are allotted sections of track on its surface as digit stores. Translation information is provided on other tracks on the same drum. Dialled pulses are stored in binary code in their appropriate positions as small areas of magnetization on the drum surface. Common electronic circuits working on a time-division system synchronized with the drum are used to 'write in' digits, refer these digits to the translation 'library', transfer the translation to its appropriate register track section and transmit this information as Strowger pulses. The drum, together with its common control circuits, provides 48 registers complete with translation facilities. A particular advantage of this technique lies in the absence of a conventional translation field. Translations can be set up or changed by a simple keying procedure.

The equivalent electro-mechanical register-translator for director areas employs binary digital storage on relays and relay counting circuits, whilst its non-director counterpart provides a further comparison in techniques by using uniselector storage and counting. Both equipments use relay translators common to 15 registers. Registers requiring translations apply to the translator and are connected in turn by relay switching under control of a uniselector. Translation digits are returned to the

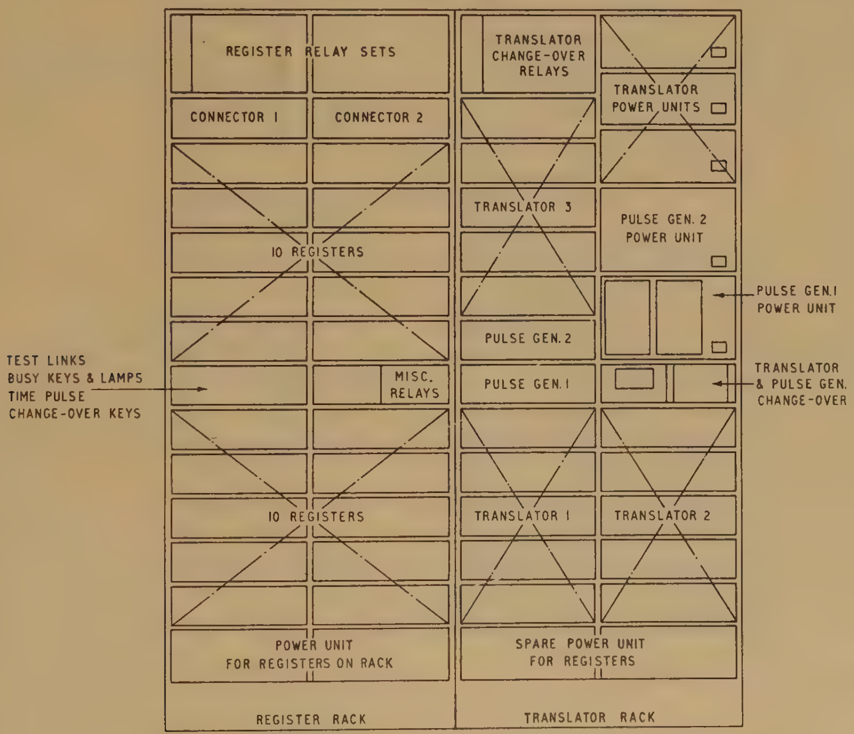


Fig. 4.—Electronic registers and translators: rack layout.

The register rack comprises:
Register relay sets.
Two connectors.
20 registers.
Register power unit.

The translator rack comprises:
Three translators (one standby)
Two pulse generaters (one standby).
Spare register power unit.
(A translator rack can serve 4 racks of 20 registers each = 80 registers.)

Size of racks: 10 ft $6\frac{1}{2}$ in × 4 ft 6 in.

Fig. 4 shows a typical arrangement of registers and translators on a rack, and Fig. 5 shows details of the register construction, as used at Bristol.

The main register-translator functions are the same for director as for non-director areas, but detailed requirements differ. In register one at a time as required, the translation time being about 150 millisec per digit.

Detection of the end of dialling has to be catered for in the register design, since a national number may contain 8, 9 or 10 digits. Positive identification by code examination would be

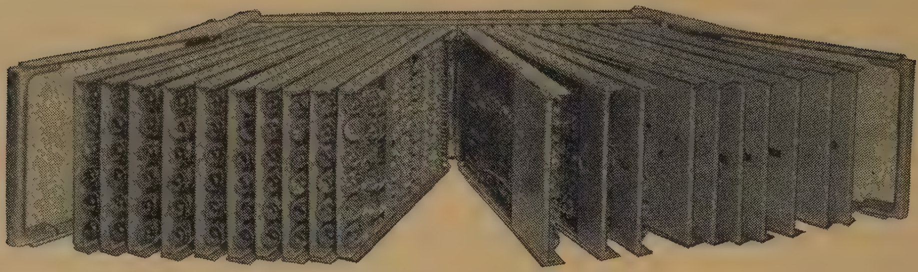


Fig. 5.—Register removed from rack.
Showing 'leaf' construction.

possible but expensive, and a simpler system based on observations of subscribers' dialling habits has been adopted. The register waits 4 sec after receipt of the 8th digit, and if a 9th digit has not then been received it assumes that dialling is complete. However, if a 9th digit is received within this period the procedure is repeated. Flexibility to vary the waiting time is incorporated in case experience should prove this to be desirable.

(4.3) Centralization of Register-Translators

The use of a prefix digit to gain access to the register-translators permits them to be located at a point remote from the originating exchange, and as trunk traffic often represents only a small percentage of the total originating traffic the concentration of the register-translators at a central point offers considerable economic advantage. However, as the register-translators are also to be used for charging purposes, it is necessary for signals to be sent back to the originating exchange to control the operation of the calling subscriber's meter. This could be done either by indicating to the originating exchange the appropriate charge rate and then generating the meter pulses locally, or alternatively by sending meter pulses back over the junction during conversation. The latter scheme is preferred, as it involves the generation and control of meter pulses at fewer points. It requires the use of a suitable signalling system to enable pulses to be passed over the junction during conversation. Such a system has been devised and is being adopted by the Post Office.

The general picture, therefore, is of a group of exchanges working to controlling registers at a suitable switching centre, usually the principal exchange in the charging group. Certain small groups will, however, have their calls switched at the centre of an adjacent group, at which the register equipment for such small dependent groups will be located. The register-translators control the setting-up of the call and determine the charge rate, while other associated equipment enables meter pulses to be passed back over the junction to the originating exchange as the call proceeds.

(4.4) Trunking Arrangements

The trunking arrangements for the s.t.d. equipment vary somewhat with the type of exchange, and three typical arrangements for giving access to the register-translators are shown in Fig. 6.

Fig. 6(*a*) shows a non-director main exchange with the '0' level trunked direct to register-access relay sets which are included in the speech path and make connection with the register-translator while the call is being set up. After transmission of the subscriber's number, the register-translator is released but the access relay set remains in circuit. It is from this relay set that periodic pulse metering is controlled.

Fig. 6(*b*) shows a non-director or satellite exchange with the '0' level trunked over special junctions to register-access relay sets at the main exchange. This arrangement is somewhat

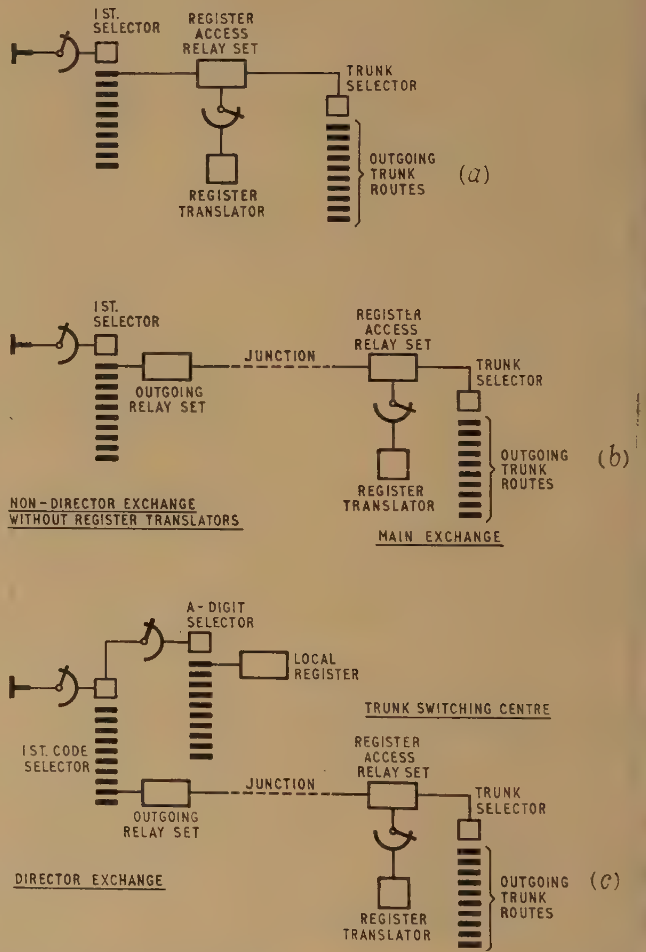


Fig. 6.—Typical trunking arrangements showing access to register equipment.

(*a*) Non-director main exchange.
(*b*) Non-director exchange using register equipment at main exchange.
(*c*) Director exchange using register equipment at trunk switching centre.

similar to that of Fig. 6(*a*), but it differs in that connection over a junction is involved. To allow the register-access relay set to control metering at the originating exchange, it is arranged to send meter pulses back over the junction during conversation. These pulses are detected by the outgoing relay set and are converted into a form suitable for operating subscribers' meters at the exchange concerned.

Fig. 6(*c*) shows a director exchange with level '0' of the A-digit selectors trunked to a local register, and it also illustrates the access provided via junctions to the register-translator equipment at a trunk switching centre. This arrangement also

employs metering over the junction circuits. The local registers provided at each director exchange are designed to transmit a predetermined digit which positions the first code selector on the corresponding level. This level gives access to the junctions to the centralized register-translator equipment. Following the transmission of this single digit, the local register is merely required to accept the digits which the subscriber dials and repeat them forward to the register-translator at the central point.

Small exchanges remote from the switching centre will employ trunking arrangements similar to those shown in Fig. 6(*b*) but modified to permit all classes of traffic—s.t.d., local and manual-board—to be routed over a common group of junctions.

Typical trunking arrangements at a group switching centre for incoming and through traffic are shown in Fig. 7; they refer to the group area shown in Fig. 2. When the code 452 is dialled at a distant exchange, the register-translator routes the call over the trunk network to incoming first selectors at the group switching centre. The next one or two digits route the call to the particular exchange required, while the remaining digits select the required subscriber. Exchanges included within a linked numbering scheme are identified by the first digits of the subscriber's local number, so that, in the example shown, the exchange code for all exchanges in the linked numbering scheme is the same as the group code, i.e. 452.

Access to other charging groups which are normally routed through the group switching centre is given via levels '1' or '0' of the incoming selector, and it is arranged that the register-translators at the originating exchange will not permit a national call to be completed where the digit '1' or '0' is dialled after the group code. By this means, irregular routing of calls which may involve different charges is prevented. When the correct group code is dialled, however, through access is obtained by the inclusion of the digits '1' or '0' in the appropriate translation.

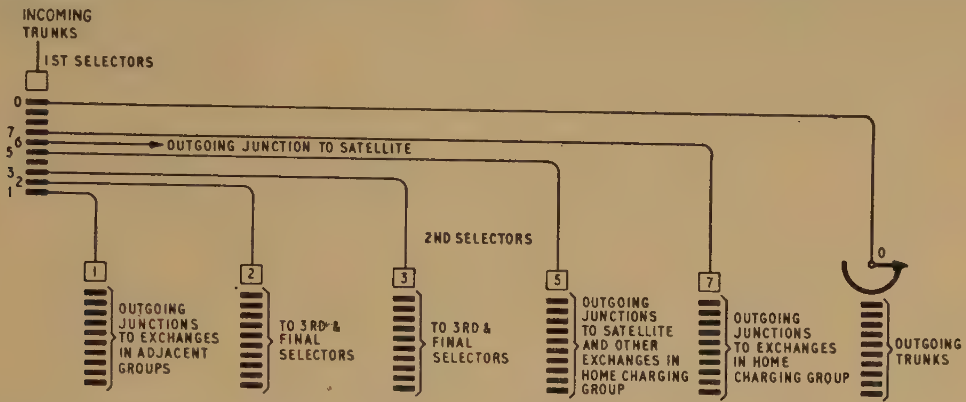


Fig. 7.—Typical trunking at a group switching centre for incoming and through traffic.

(4.5) Maintenance Features

The performance of automatic switching equipment in service rests, in the first instance, on the achievement of sound and reliable design. The British Post Office and the manufacturing industry in Great Britain have, throughout the years, devoted a large part of their effort to designing for reliability. The steady and progressive process of applying, in design, information obtained from laboratory and field trials, and from service experience, has resulted in equipment of such reliability that a high standard of service can be maintained with low maintenance effort.

Nevertheless, however good the design, some faults of a wear and tear and random nature will inevitably occur in the complex network of apparatus and lines forming a fully automatic national telephone network. The equipment has been designed to minimize the service effect of any such faults. For example, it is arranged, as far as possible, that a subscriber making a second-attempt call will use different equipment, and common equipment will, in general, be duplicated, with automatic change-over and alarm facilities actuated from self-checking and monitoring circuits.

(4.6) Further Development Work Now in Hand

It is proposed to provide fast switching equipments at transit exchanges, with rapid signalling between them. The existing trunk routes, equipped with dial pulsing systems, will be used for terminal calls between directly connected areas. The new equipment, in a separate network, will cater for the rapid establishment of transit calls. As the majority of calls are between directly connected areas, not much new plant will be required.

The transit network is being planned on the basis of 4-wire switching, with facilities for automatic alternative routing, and with 4-wire transmission paths. The use of the proposed fast switching and signalling equipment will reduce the maximum delay in receipt of ring tone to the order of 7 or 8 sec. It is probable that the fast signalling will be in coded form, and this will enable features to be incorporated to detect errors occurring during numerical transmission. It is also intended that the line signalling systems for the new network should be designed to take circuits out of service automatically during times that fault conditions make the transmission paths unusable. These arrangements should contribute to the attainment of a high quality of service on all calls.

(5) TARIFFS

(5.1) The Call Charges at Exchanges not Provided with S.T.D. Facilities

Since the introduction of group charging on 1st January, 1958, the charge for local calls from subscribers—except those with coin boxes—has been 3d.; the fee from a coin-box telephone is 4d. Such calls are untimed. The local-call charge applies to all calls within the local fee area, which, on the average, covers about 900 square miles.

Trunk calls, which are set up by operators, are timed on the basis of an initial charge for the first three minutes of conversation and proportionally for each subsequent minute. The trunk-

call charge varies with the distance measured between group measuring points in the following manner:

Up to 35 miles	1s. for 3 min.
From 35 to 50 miles	1s. 9d. for 3 min.
From 50 to 75 miles	2s. 3d. for 3 min.
From 75 to 125 miles	3s. for 3 min.
Over 125 miles	3s. 6d. for 3 min.

During the cheap-rate periods, the charges for calls above 35 miles are reduced to approximately two-thirds of these rates. From 1st July, 1958, the cheap-rate periods have been:

Monday to Saturday (inclusive) ..	6 p.m. to 6 a.m. next day
Sundays	2 p.m. to 6 a.m. Monday

The local-call charge, and the charge for trunk calls up to 35 miles remain the same day and night. During both the full- and the cheap-rate periods, an additional charge of 3d. is made for a trunk call from a coin-box telephone. The charge for a local call is registered on a meter associated with the calling subscriber's line, but subscribers' trunk-call charges are calculated from details entered on a ticket by the controlling operator at the time the call is made.

(5.2) The New S.T.D. Call Charges

The choice of a suitable tariff for subscriber trunk dialling was a matter of major importance.

If it were required to perpetuate under fully automatic conditions the recording of full details of trunk calls, it would be necessary to employ some system of automatic ticketing with calling-subscriber identification to replace the manual ticketing now done by the operators. Such systems are used by some administrations, notably in America and Belgium. In most European countries, however, the need for complex and relatively expensive automatic ticketing equipment has been avoided by making greater use of the subscribers' meters already existing, and this principle is to be used in the United Kingdom.

The system of charging to be used for subscriber-dialled trunk calls is known as 'periodic metering' and involves the registration of a single unit fee at intervals throughout the duration of the call. The interval varies with the chargeable distance appropriate to the call in question. There is some further simplification in that the number of distance steps for trunk calls is reduced to three. This reduction in the number of charge steps is coupled with a reduction in the unit fee itself, and on the introduction of the new system the unit fee for subscribers except those with coin boxes will become 2d. For calls from coin-box telephones it will be 3d. The new trunk tariff for subscriber-dialled calls can be expressed in the following way:

	Ordinary lines	*Coin-box lines*
Up to 35 miles	30 sec for 2d.	30 sec for 3d.
From 35 to 50 miles ..	20 sec for 2d.	20 sec for 3d.
Over 50 miles	12 sec for 2d.	12 sec for 3d.

In the cheap-rate period (as defined in Section 5.1) the unit fee will, on trunk calls, buy half as much time again.

To provide for this tariff the register-access relay set which i[s] carrying the call in question is arranged to send one meter puls[e] immediately the call is answered, to operate the calling sub-scriber's meter and, in the case of coin-box lines, to control th[e] application of pay tone (see Section 7). For other than coin-bo[x] lines, this pulse in effect charges for the first interval o[f] time, whatever that may be. The relay set is also designed t[o] receive from the register-translator an indication of the particula[r] periodic metering rate applicable, and on a trunk call it will hav[e] preselected one of three pulse supplies in readiness for metering. Each pulse supply has a frequency six times that of the corre-sponding metering rate, and whenever an answer signal i[s] received, a mechanism in the relay set starts to count thes[e] supply pulses. This mechanism is so arranged that it first send[s] out a meter pulse when the seventh supply pulse is received and thereafter sends a meter pulse for every six supply pulse[s] received. The effect of this is illustrated in Fig. 8, and it will b[e]

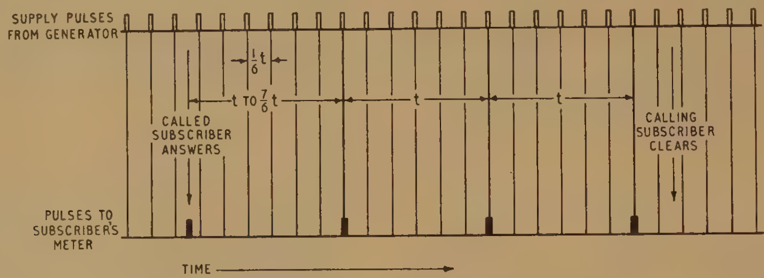


Fig. 8.—The pulse-metering principle adopted.

seen that the subscriber is safeguarded against any loss of pai[d] time which might arise if the metering pulses were of rando[m] incidence.

The supply pulses are provided by a rotary machine which i[s] capable of supplying 20 different rates. Only four of these a[re] connected to the register-access relay sets at any one time, thereb[y] giving the three trunk rates mentioned, plus a local call ra[te] which will be referred to later. To cater for the cheap-ra[te] periods, a different selection of four pulse rates must be mad[e] from the 20 that are available, and to meet this requiremen[t] automatic reselection at a predetermined hour of the day is pro-vided by control equipment interposed between the pulse gene-rator and the relay sets. This control equipment also provid[es] for monitoring of the pulse supplies and automatic change-ov[er] of generators at regular intervals or under fault conditions. Th[e] general arrangement of equipment is illustrated in Fig. 9.

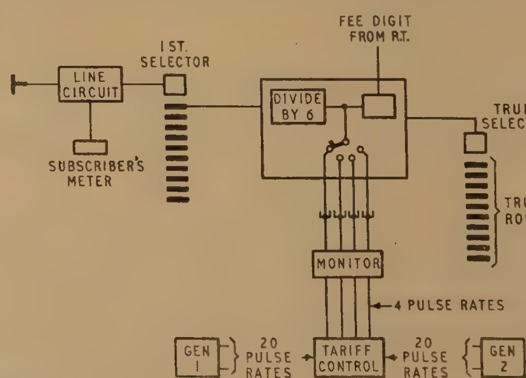


Fig. 9.—General arrangement of pulse-metering equipment.

The system of periodic metering adopted for this count[ry] has advantages both for the subscriber and the administratio[n]. The considerable advantage which subscribers gain is th[e]

s.t.d. trunk calls will not be subject to the present 3 min minimum charge. A trunk call of sufficiently short duration can, in fact, be made for as little as 2d. There is a similar advantage if a call runs just over three minutes. Instead of an immediate increase to cover one additional minute, the charge will increase by no more than 2d. for each 12 sec, even at the maximum trunk rate. The advantages which the administration gains from this system result from its simplicity. Existing subscribers' meters are used to record the unit charges; it is unnecessary to provide equipment to time each call individually,

Fig. 10.—Charging rates from Bristol Central Exchange under subscriber trunk dialling.

although the system gives a close approximation to the results obtainable with individual timing; by avoiding complex arrangements maintenance is made easier. A periodic metering system of this nature thus limits expenditure upon equipment both initially and as a recurring charge, and reduces accounting costs. It therefore assists the administration to operate at low cost, and ultimately helps to reduce the subscribers' service charges.

It has been mentioned that the local-call area was, on 1st January, 1958, extended to cover the home and adjacent charging groups, the charge being 3d., untimed, for calls from an ordinary telephone, and 4d., untimed, for calls from a coin-box telephone. It would not be possible to reduce these fees unless timing were introduced, and this, in fact, will be done in s.t.d. areas. On local calls, the unit fee of 2d., for calls from an ordinary telephone, or 3d., for calls from a coin-box telephone, will buy 3 min time. In both cases, the unit fee will buy twice as much time, i.e. 6 min, on local calls made during the cheap-rate periods. The great majority of local calls do not involve the use of a register-translator or its associated access relay set, but are completed via other paths through the exchange switching equipment. Small relay sets designed specially for the purpose of timing local calls will be connected in these paths. When a call is answered the normal exchange equipment is used to supply a single meter pulse which is registered on the calling subscriber's meter and also acts as a start signal to the local-call timing relay set. The local-call timing relay set then starts to count supply pulses from a clock-controlled source in a manner somewhat similar to that previously described when dealing with trunk calls. In this case, however, the supply pulses are run at ten times the nominal metering rate. Basically, supply pulses for local-call timing are generated by a pendulum clock, but a unit is interposed which regenerates the pulses at five different rates to meet long-term requirements, provides for doubling the interval between pulses to provide the cheap rate, and delivers the final supply pulses to the relay sets in nine different phases to avoid overloading common battery supplies.

Clearly the introduction of subscriber trunk dialling must be a gradual process, and the new tariff and the old will be co-existent for a long time. The new tariff will apply to all local and subscriber-dialled trunk calls at exchanges equipped for subscriber trunk dialling, and subscribers on these exchanges will thus have the benefit of cheap short-duration trunk calls. The effect of the application of the new tariff to calls originated in the Bristol Central Exchange Area after the opening of the s.t.d. service there is shown in Fig. 10. The old tariff will continue to apply to trunk calls which are set up by an operator at s.t.d. exchanges and to all calls at exchanges not equipped for subscriber trunk dialling.

The adoption of periodic metering for both local and trunk calls in the new tariff will give considerable tariff flexibility, and will enable adjustments to be made in the time purchased by the unit fee for any particular class of call, without affecting other charges.

The facility of making a short-duration trunk call for the unit fee will also provide an alternative to the personal-call and transfer-charge services. A subscriber will, instead, be able to call the distant number at a small charge to find if the required person is there, or to ask him to establish a call in the reverse direction.

(6) ACCOUNTING

(6.1) The Present Arrangements

Telephone bills for subscribers without subscriber trunk dialling show, besides rental and other charges not relating to calls, a total charge for local calls, whether metered or ticketed, and a total charge for trunk calls. In addition, a statement is provided showing the date of, and charge for, each trunk call. Extra details about individual trunk calls are supplied on request at an extra charge, either on the trunk-call statement or at the time the call is made. Normally subscribers are billed half-yearly, but many large users are billed monthly for their trunk calls.

(6.2) The S.T.D. Arrangements

For subscriber trunk dialling, the arrangements have been recast to fit the new conditions. As dialled trunk calls will be recorded at the exchange on the same meters as local calls, separate information about dialled trunk calls will not normally be available, and a bulked bill will be unavoidable. The frequency of billing of s.t.d. subscribers will be quarterly instead of half-yearly. Each quarterly bill will show the number of units and the total charge for metered calls, as well as separate totals of charges for local and trunk calls made via an operator. A detailed statement giving the date and charge for each trunk call will only be given for calls made via an operator.

Fig. 11.—Subscribers' private meters.

(6.2.1) **New Facilities.**

Certain new facilities will be available. For some subscribers' installations, such as those at hotels and clubs, information about the charge for individual calls, at the time they are made, is often essential. S.T.D. subscribers will, of course, still be able to obtain details of the charge for individual calls by passing them via an operator even if they could have dialled the calls themselves. However, since the charge for calls via the operator will be at the old tariff and therefore usually higher, this would not provide a satisfactory general solution. Also some subscribers, particularly larger users, like to keep a watch on their trunk calls. Meters have therefore been designed for fitting to telephone installations at trunk-dialling subscribers' premises, on a rental basis, to indicate both the number of chargeable units for individual calls and a cumulative total. These 'private' meters are of two kinds—a clock type of meter for use with individual telephones and small p.b.x.'s, and a cyclometer type of meter for use with larger p.b.x.'s.

The meters will operate in step with the corresponding meters at the exchange. Each time a meter pulse is recorded on the exchange meter a 50 c/s pulse of current is applied via an injection transformer longitudinally to the subscriber's line to operate the private meter.

Two versions of the clock-type meter, differing in appearance but providing the same facilities, have been developed and are shown in Fig. 11. Provision is made for showing the total units charged up to 9999, and on a separate resettable hand the units charged on individual calls up to 99.

The cyclometer-type meters will be operated by small metering units which receive the a.c. pulses from line and convert them to direct current. For individual call-charging resettable meters will be provided, either permanently associated with one line or with switching to any one of ten lines. Apparatus has also been developed for use at the exchange to enable an automatic record to be made of the digits dialled and the meter pulses recorded on any subscriber's line, where this may be found necessary for service reasons.

(6.3) The Preparation of Accounts. Development Trends

The work involved in the preparation of accounts is considerable. It involves nearly 10 million meter readings and over 700 million tickets each year. Subscribers' meters are, at present, mostly read by operating staff. Other methods of obtaining meter readings are being developed. These include photography of existing types of meters for use in connection with punched-card processing, and an electronic meter-pulse recording system using magnetic drums and tapes. In conjunction with the latter, the possibilities of a system of automatic data-processing for the production of subscribers' bills, in which the exchange equipment will be closely integrated with the processing system, is also under examination. Such a system would be co-ordinated with the methods of processing tickets for manually-handled calls, using mark-scanning and punched-card techniques, which are already being extended throughout the country.

(7) COIN-BOX SERVICE UNDER SUBSCRIBER TRUNK DIALLING

To enable call-office users to dial their own trunk calls, a new coin box has been developed using a 'pay on answer' system. The design has been based on the principle that the box itself should be as simple and inexpensive as possible, the controlling facilities being incorporated in the associated equipment at the exchange. Control by button A and button B, as used on the existing coin boxes, has been eliminated. A general view of the box is shown in Fig. 12, and the main elements are shown in Fig. 13.

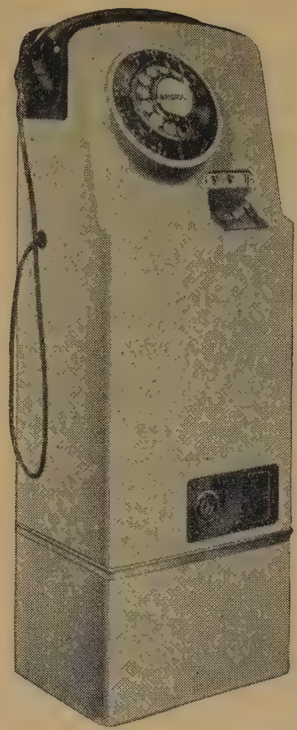


Fig. 12.—The 'pay on answer' coin box.

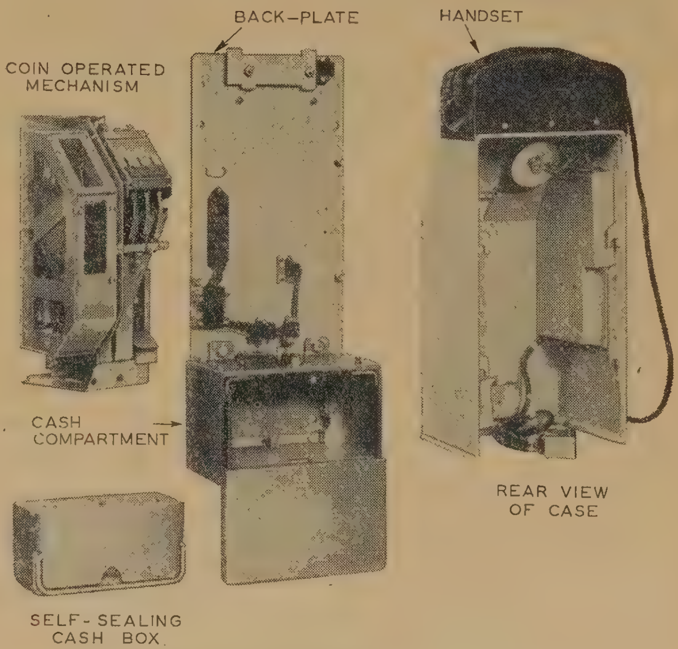


Fig. 13.—Main elements of the 'pay on answer' coin box.

A basic feature is that a caller dials the required number *before* inserting any coins, and is, in fact, prevented from inserting coins at this stage by a locking bar on the coin slots. When the called subscriber answers, both parties hear an interrupted tone ('pay tone') and the coin slots are unlocked. The caller must now insert a coin (3d., 6d. or 1s.) during the next ten seconds to establish conversation. When the time paid for has expired the caller is advised by a brief reapplication of pay tone. A

short period is allowed for inserting coins or for finishing the conversation, and if further coins have not then been inserted the call is terminated.

Loop-resistance pulses are used to signal the value of the coins to the exchange, where these pulses are recognized, stored, and compared with the meter pulses to determine when pay tone should be reapplied. Coin pulses are generated by cam-operated spring sets in the coin-operated mechanism. The cams are raised as the coin is inserted, and when the coin drops inwards the cams commence to fall at a governed speed. The first half of their movement allows time for the coin to be tested, and during the second half, coin pulses are generated and transmitted to line, provided that the coin has been accepted. On manually controlled calls the operator controls the opening of the slots and hears tone signals for each coin inserted. As refund of money is not provided for, provision has been made for the operator to obtain signals to enable her to check the amount deposited at any time.

(8) THE ARRANGEMENTS FOR INTERNATIONAL SUBSCRIBER DIALLING (I.S.D.)

(8.1) The Present Position of the International Telephone Service

The position reached in the mechanization of the international telephone service is that signalling and switching specifications have been established through the C.C.I.T.T. with a view to enabling an operator in the outgoing country to establish a call automatically without the assistance of an operator in the incoming country. Equipment conforming to these specifications is being installed in a number of international centres throughout Europe, including London, and will commence to come into service towards the end of 1958. Although the C.C.I.T.T. specifications were prepared primarily for semi-automatic operation, they are basically suitable for fully automatic working. There already exist certain i.s.d. installations of a special and restricted nature set up between countries by bilateral agreement.

(8.2) International Accounting

Among the many problems which arise with fully automatic operation is the question of the methods to be adopted for the settlement of international accounts between European countries. In the international telephone service the revenue obtained from a call is divided, in agreed proportions, between the outgoing country, the incoming country and any transit country through which the call may be routed. With semi-automatic operation the necessary accounting information, e.g. the destination of the call, the route followed and the chargeable duration, can be extracted from call tickets prepared by operators. With fully automatic operation this source of information is no longer available, and if international accounting is to continue to be based on the actual amount of traffic exchanged and the route followed between countries, the information required for the preparation of international accounts will need to be extracted at some convenient point in the network and recorded automatically. The practicability of doing this is at present under intensive study by the C.C.I.T.T.

(8.3) Register-Translators for International Calls

In the United Kingdom, international subscriber-dialled calls will be handled by register-translators situated at group switching centres, and a 3-digit access code has been allocated for this purpose. Although it is the intention to route all international calls via the London Continental automatic exchange, these register-translators will be required to determine, from the number dialled, the charge rate to be applied to the subscriber's meter. This will involve an examination of the country code (two digits), and also, where there is more than one charge zone in the incoming country, one or two digits of the national number of the called subscriber.

(9) CONCLUSION

It is confidently hoped that the progressive introduction of a subscriber-trunk-dialling scheme on the basis described in the paper will further encourage the rapid, economic and efficient development of the telephone service in the United Kingdom. The arrangements made envisage the later introduction of fully automatic international telephone operation.

The successful development of the necessary new telephone-exchange switching and signalling equipment, and the manufacture and installation of the Bristol equipment within programmed dates, owe much to the excellent co-operation which exists between the Post Office and the leading manufacturers of telephone-exchange plant in the United Kingdom.

(10) ACKNOWLEDGMENTS

Acknowledgment is made to the Engineer-in-Chief of the Post Office for permission to make use of the information contained in the paper. The author also wishes to thank a number of his colleagues for their suggestions and assistance in the preparation of diagrams.

(11) BIBLIOGRAPHY

(1) RADLEY, W. G., and WRIGHT, E. P. G.: 'Voice-Frequency Signalling and Dialling in Long-Distance Telephony', *Journal I.E.E.*, 1942, **89**, Part III, p. 43.

(2) STERKY, H.: 'Telecommunications in Sweden—Present and Future', *Tele. Information from the Swedish Board of Telegraphs*, 1950, No. 1.

(3) LAWRENCE, J. A.: 'Contemporary Telephone Mechanisation Abroad and Future Trends', *Post Office Telecommunications Journal*, 1950, **2**, p. 154, and 1951, **3**, p. 53.

(4) THOMPSON, A. J., and BROADHURST, J. H.: 'Trunk Mechanisation' (Institution of Post Office Electrical Engineers, Printed Paper 201, 1951).

(5) GIBSON, R. W.: 'National Trunk Dialling', *Post Office Electrical Engineers' Journal*, 1951, **43**, p. 169.

(6) BARRON, D. A.: 'Subscriber Trunk Dialling in the United Kingdom. A Review of the Problems' (Institution of Post Office Electrical Engineers, Printed Paper 203, 1952).

(7) CLARKE, A. B., and OSBORNE, H. S.: 'Automatic Switching for Nationwide Telephone Service', *Bell System Technical Journal*, 1952, **31**, p. 823.

(8) PILLIOD, J. J.: 'Fundamental Plans for Toll Telephone Plant', *ibid.*, 1952, **31**, p. 832.

(9) NUNN, W. H.: 'Nationwide Numbering Plan', *ibid.*, 1952, **31**, p. 851.

(10) SHIPLEY, F. F.: 'Automatic Toll Switching Systems', *ibid.*, 1952, **31**, p. 860.

(11) 'Notes on Nationwide Dialling' (American Telephone and Telegraph Co., 1955).

(12) C.C.I.F. XVIIth Plenary Assembly Vol. V, 1955.

(13) 'Notes on Distance Dialing' (American Telephone and Telegraph Co., 1956).

(14) 'Full Automation of the Telephone System', Government White Paper, Cmnd. 303, 1957.

(15) 'Telephone Policy. The Next Steps', Government White Paper, Cmnd. 436, 1958.

DISCUSSION ON THE ABOVE PAPER

Before THE INSTITUTION 22*nd January, the* SOUTH MIDLAND RADIO AND MEASUREMENT GROUP *at* BIRMINGHAM 26*th January, the* NORTH-EASTERN RADIO AND MEASUREMENT GROUP *at* NEWCASTLE UPON TYNE 2*nd February, and the* NORTH-WESTERN CENTRE *at* MANCHESTER 3*rd March*, 1959.

Sir Gordon Radley (*at London*): The author draws attention to the fact that most countries well developed telephonically have introduced, or are planning for, subscriber trunk dialling in a form suited to their requirements. They may be divided into two broad groups. First, mechanization of the trunk and local services is virtually complete and showing very considerable operating savings in countries such as The Netherlands and Switzerland. But, these countries do not have as many telephones as there are in Greater London, where a linked numbering scheme has given interdialling facilities to about 2 million telephones for a long time. Second, some subscribers in the United States have been able to dial nationwide calls for some years. But many of the other features which distinguish a trunk from a local call remain—in particular, the arrangements for making an individual record of each call. This results partly from State and Federal requirements, but it has necessitated the installation of a great deal of elaborate and costly equipment, and sometimes the retention of operators.

The Post Office took the deliberate policy decision to simplify arrangements and to use the minimum amount of elaborate equipment, although this meant abandoning all provision for the detailed charging of trunk calls. The aim, in very broad terms, is to convert the telephone system of this country into one vast local network in which local and long-distance calls will be timed and metered in exactly the same way; the only difference will be that, for a long-distance call, meter operations will take place more frequently than for a local call.

During the last 20 years operators' wages have increased about three times. The cost of providing a long-distance circuit has not increased to anything like that extent. But long-distance circuits can be provided in modern h.f. systems at a cost less than that of the corresponding a.f. circuits of the pre-war era. The greater the capacity of these high-frequency systems becomes the lower will be the cost of an individual circuit in them and the greater the disparity between the cost of the plant that is used and the cost of any manual operation and accounting. Furthermore, with the modern high-frequency system nearly all the plant costs are concentrated at the ends. Roughly three-quarters of the cost of a 100-mile coaxial cable equipped to provide 960 circuits is in the channel, signalling and power equipments at the ends; only one quarter is in the line. Within Great Britain the cost of a trunk circuit will not be proportional to its length. With subscriber trunk dialling all calls over 50 miles are charged the same amount. Ultimately all trunk calls may be charged at one rate.

Group charging introduced considerable simplifications about a year ago. In particular, the decision that calls to adjacent groups should be charged at the same rate as 'within-group' calls did away with the necessity for quite a lot of equipment. As the Postmaster General said, technical developments have been facilitated by administrative simplification.

The performance of the electro-mechanical and electronic register-translators will be watched over the next few years. Functionally, both are quite adequate to the job. But the speed of some of our present switching equipments may be inadequate. I was very glad to note that the use of fast switching equipments will reduce the maximum delay in receipt of ring tone to 7 or 8 sec.

Plans have been announced that, by 1970, three-quarters of all trunk calls should be dialled by the caller. There may be pressure to accelerate the rate of automation. Probably a year or two could be knocked off the programme. Neither the rate of investment nor manufacturing capacity is likely to be the limiting factor. We now have over 50 000 telephone operators, and we may only want about 20 000. The rate at which we can redeploy the remainder may very well have much bearing on the timetable.

Mr. F. O. Morrell (*at London*): The paper illustrates two problems of telephone engineering. First, facilities must be provided as economically as possible, and second, the telephone engineer faces the problem of 'grafting' them on to a vast mass of existing equipment. He must therefore look as far as possible into the future without making the present unduly expensive, and the result is a compromise, of which there is evidence in the paper.

Let us consider, for example, the numbering plan. A universal national numbering scheme is not practicable, and eight, nine or ten digits will be used, introducing a 4 sec delay in many cases. This is a long time compared with the 7 sec quoted in Section 4.6 of the paper. A more expensive register-translator could identify the code and obviate the delay, and the particular arrangement described has been adopted as a compromise between cost and facilities.

Another example is the 'own-exchange-only' type of register-translator. It is less expensive than the 'right-through' type, but it could lack flexibility, and may not be able to provide, say, alternative routing on a wide scale. This may not be vitally important in this country, which, in number of telephones per square mile, is so densely populated that the circuits between two centres can be spread over more than one route, but if the system expands to cover the whole continent of Europe, and later, the world, alternative routing and re-routing may have to be provided on a wide scale.

With both examples in mind, one would perhaps like to envisage a register-translator in the local centre which would scrutinize the number dialled, and, interpreting the whole code, survey the appropriate part of the network and route the call in the best possible way.

I think that this could be done, even at the present time. Electronic register-translators are capable of providing virtually unlimited facilities. They would be more expensive than necessary in meeting present requirements, but the future telephone engineer may wish they had been provided from the beginning. I should like to have the author's views on this matter.

Mr. L. S. Crutch (*at London*): The paper reveals the position of the subscriber-trunk-dialling service in relation to similar facilities in other countries, and forecasts the position in this country by 1970. One could wish that more rapid progress were possible, both for the service itself and for the prestige position in world markets. The author is concerned with the technical aspects of the situation, but it is well known that other factors can influence the rate of provision of the service.

The ability to give a subscriber a trunk-dialling service arises from many developments in the telephone art over the last 25 years. I believe that the main contributions have come from voice-frequency signalling systems and from multi-outlet fast-searching switches. The former was originally introduced to improve manual supervision, later extended to set switches under the control of an operator's dial, and is now being developed into something like a data-sending system. The latter has been used extensively for trunk switching.

We must also not overlook the developments in transmission techniques which have made possible the economical provision of circuits of high quality on a large scale. We have had a demonstration showing that a caller can dial 8 or 10 digits, and may

reach the wanted subscriber in less than 10 sec. One has only to recall the procedure and delay in making trunk calls 25 years ago to realize what an immense improvement in service is now being described.

The paper suggests that the new service will give rise to a gradual reduction in the total number of operators. I believe that this view was held in the United States, but they are now finding that the greatly increased use of the service and the need to retain operators in some circumstances reduces the fall in numbers which had been contemplated.

With regard to the new coin box, it appears that you are unable to talk to your wanted subscriber during the 'pay now' period. This suggests the need to educate private subscribers into the habit of announcing their number or identity instead of the practice of answering a telephone call with the word 'Hello'.

Mr. F. I. Ray (*at London*): Automation can occur in two ways: by using a machine to do precisely what is being done manually or by altering the conditions to suit the machine. We have chosen the latter. In subscriber trunk dialling we have not invented a machine which does precisely what the operator does. We have altered our tariffs to suit the capability of the machine.

My second point relates to the principles listed in the paper. These include a provision that the system should be easy for subscribers to understand and use. How are we to know what is easy for a subscriber to understand? Is it easier to use a system with a uniform number of digits, or one in which the subscribers dial the minimum number of digits? Both can claim to be simple and easily understood.

The observation results from Bristol are now to hand. They are taken on a comparatively small sample and the figures may alter when we get more results, but they are extraordinarily satisfactory. 96% of the calls which can be dialled are being dialled. That exceeds our highest expectations and shows the value of giving subscribers an incentive to dial. The percentage of failures compares favourably with those obtained when operators dial long-distance calls. We have introduced no appreciable deterioration in service with our electronic s.t.d. equipment. The equipment is additional to that for a local call, and the reliability of subscriber trunk dialling depends more than anything else on the standard of maintenance of the local exchanges.

We expected dialled trunk calls to be more numerous but shorter. The statistics confirm this, but the number of calls has increased by 40% and the durations are down by only 20%; the the traffic in erlangs has increased. And the subscribers seem to be very satisfied with the new service.

Mr. C. Riley (*at London*): Post Office engineers have had to 'graft' the s.t.d. system on to a vast network of existing exchanges which must necessarily form part of the final pattern. The fact that this has been found possible without serious technical difficulties illustrates the great flexibility of the basic step-by-step Strowger system, which was standardized so many years ago by the Post Office.

An s.t.d. system depends for its attractiveness on having a high percentage of subscribers in the country connected to automatic exchanges, and it is disappointing that this country is next to the bottom amongst those countries with a high telephone development.

The problem of the time interval between the end of dialling and receipt of the ringing tone is quite a serious one. If this is substantially greater than the 7 or 8 sec quoted, would it not be possible to introduce high-speed signalling on the existing trunk routes?

The charging-area diagram shows the great simplification introduced, but the grading does seem somewhat coarse, since there is only one grade outside the 50-mile area. Thus, in spite of all the trouble which has been taken to eliminate or reduce anomalies on the boundaries of charging areas, a call from, say, Bristol to Exeter will be charged at the same rate as one from Bristol to Glasgow or further. Figs. 8 and 9 show that technically it would have been easy to obtain finer grading.

The new coin box is a very clean streamlined design, but there is nowhere to place money before insertion. In view of the short interval permitted before a call is lost unless further coins are added, I suggest that suitable provisions be made in the design of the kiosk.

Mr. J. R. Pollard (*at London*): I would like to refer to apparatus under development which has a number of operating advantages over the systems described. The equipment makes extensive use of time-sharing principles, as used in some kinds of electronic telephone exchanges, and an outline of its operation is shown in Fig. A. Each register consists of a column of cores. Storage is

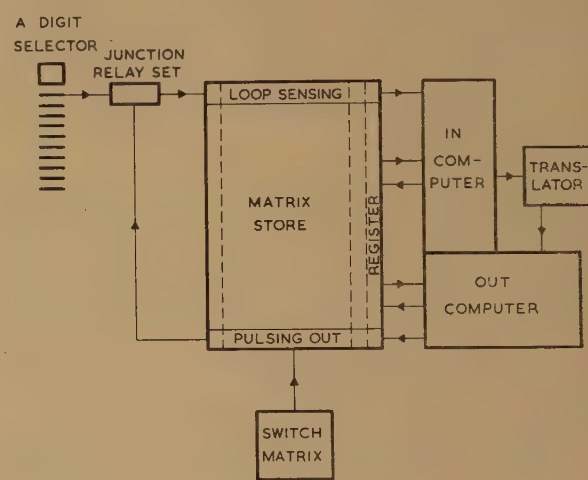


Fig. A

provided, by means of these cores, for the state of the incoming loop, the state of the outgoing loop, the code and numerical digits dialled in, the translation digits and the fee.

In operation, each register is read out in turn and the state of the input loop is compared with that stored. The sampling process is fast enough to enable impulses to be counted and routed into successive digit stores as each number is dialled. These operations are performed by the 'in' computer, which is common to all registers. Information is also transferred, one register at a time, to the translator, where, if sufficient information is recorded in the register, a translation is produced and the required digit is recorded in the register.

To pulse out, the appropriate part of the register is read into the 'out' computer, which instructs the relay set to send a 'break' or 'make' or tones for high-speed signalling as required.

Present indications are that a typical installation of the proposed equipment would take about half the space of the cold-cathode equipment. This economy in space arises partly from the saving of equipment by using the time-sharing principle and partly by the use of ferrite cores and transistors. The power consumption is very low and no high-voltage supplies are needed. For an installation of 100 registers the total consumption would be about 150–200 watts, of which a significant proportion would be associated with the access relay sets.

Mr. J. A. C. King (*at London*): The author states that a problem arises from the fact that the number of digits which has to be dialled for a call varies from one district to another. Surely this can be very simply overcome. Once one has dialled into the long-distance selector equipment with a zero and used a 1-, 2- or

3-figure code to select the area to which one's call is going, in common with the charge indicator which is returned, some sort of indication can be sent back from the central equipment to denote whether this is to be followed by four, five or six more figures. As soon as one has selected the area, one knows how many more cyphers must follow to identify the particular called subscriber.

Mr. L. L. Tolley (*at Birmingham*): A good deal of preparatory work has been done in Birmingham as in the rest of the country. The grouping of local exchanges mentioned in Section 3.2 was brought into effect just over a year ago. The automatic trunk exchange in Birmingham came into service over a year ago. It is not yet available to subscribers, but operators are dialling through it, and it is in circuit on many of our trunk calls. Birmingham has a director system of local exchanges, and therefore its register-translators will be somewhat different from those at Bristol as mentioned in Section 4.2. These register-translators and the equipment for the local exchanges are in the process of being ordered, and we expect that, in a little over two years, the equipment to serve nine exchanges, mostly towards the centre of the city, will have been manufactured and installed. Orders are also being placed to provide the service to 36 other exchanges in the Midlands.

The trunk automatic exchange for Birmingham cost £1 million. This serves the whole of the zone, including Coventry, Worcester and Wolverhampton. There is also a trunk automatic exchange at Nottingham and one is being installed at Leicester. The cost of local equipment still to be installed is estimated at £70 000 for each of the larger exchanges. As the telephone service is run on a commercial basis, obviously the Post Office expects that this very large capital investment will be repaid by increases of traffic and savings in operating costs. Could the author give a résumé of expected trends?

The provision of switches in trunk automatic exchanges has been on a more generous basis than is now adopted for local exchanges, and, in view of the greater number of switching stages in a long-distance call, I presume that the more generous basis will continue to apply to trunk exchanges. Trunk and junction circuits between exchanges are on a more restricted basis. I should like to have the author's views on the basis of provision of trunk and junction circuits in relation to subscriber trunk dialling.

The introduction of subscriber trunk dialling will tend to emphasize the effect of a casual fault. The testing arrangements for local automatic exchanges have been developed over many years. They include testers to check the working of individual items of equipment, observations by operators of the switching processes and automatic apparatus that sets up calls to selected numbers—in fact, an overall check of the exchange operation. Testers are, of course, being provided for routining the various parts of the trunk apparatus, but has consideration been given to the design of an overall tester? There would obviously be some difficulties. To avoid causing artificial congestion of the routes, the testing should be done at night, which implies that the tester should be fully automatic and capable of recording faults for investigation next day. Since the tests would be in a period of very light traffic, the tester should incorporate some means of ensuring that tests were not made only on the first- or second-choice outlets. This has to be effected in every exchange over the route being tested. The difficulties are obvious, and the tester would certainly not be cheap, but it would seem to be a most valuable tool.

Mr. N. C. Smart (*at Birmingham*): I have been concerned with the project almost from the start, and it is interesting to review the stages through which it has passed. The first was when we knew that the British Post Office was going to embark on the scheme of subscriber trunk dialling. This was particularly interesting to us, because we knew that some of our oversea customers were going to ask for this equipment, and it gave us a chance of getting experience of the problems at an early date, which has been very useful. The second stage was when we were entrusted with designing the electronic register-translator. Next we had the design approved by the Post Office, and finally, of course, there was the highlight at Bristol when it was put into service.

It is about four years since we first started on the project, and the equipment now seems rather out of date. In other fields the cold-cathode tube is obsolete. When we were entrusted with the design, the Post Office stipulated that we should use the cold-cathode tube as the counting and storing device. There were very good reasons for this, namely that when the register is functioning the glow of the tubes is clearly visible and one can count the impulses entering and being stored. The maintenance engineer is thus given a clear picture of the functioning of the register. However, the equipment is rather large and expensive, and the provision of power is difficult. Would the author use cold-cathode tubes in any new design?

Mr. J. S. Roebuck (*at Birmingham*): If you were making a long-distance call, which for some reason was interrupted, and you held on to the telephone hoping for something to happen, would you be charged for that period?

Mr. L. J. Glanfield (*at Birmingham*): The system stands or falls on the numbering scheme and its ready acceptance by the subscriber. The author states that 'One of the fundamental principles . . . is that it should be easy for a subscriber to understand and use'. He refers to the C.C.I.T. recommendation that national numbering schemes should not contain letters. Our experience has shown that, for big cities, it has proved an advantage to mix three meaningful letters into the seven digits which have to be dialled. This has been very successful, largely because the subscriber uses the meaningful system of letters on 99% of his calls. When we were studying subscribers trunk dialling in the early stages we were definitely considering 'no letters'. Much thought has since been spent on this problem. It is not only an engineering one, since the administration must play a large part. I wonder whether the system is going to be as simple for subscribers as we think. It is easy for people in a director area to regard director dialling as simple, but at Bristol we have to put a director dial on each installation.

The Bristol dialling code list shows that subscribers will use letters only on a proportion of calls, but even so, they cannot be used in a uniform manner.

We seem to have evolved a scheme which is not so simple for the subscriber as we would like, although I realize that the customer can be conditioned to anything. Is this system of dialling causing any trouble in Bristol, and, in fact, will the existence of the hybrid scheme in this country, to some extent, prejudice international subscriber dialling from the Continent?

Mr. M. S. Simcox (*at Birmingham*): Could the author state what percentage of the total trunk traffic at Bristol is being handled by the manual board?

Mr. A. E. Twycross (*at Newcastle upon Tyne*): Newcastle Central telephone exchange, one of the larger centres, is due for conversion to subscriber trunk dialling by mid-1961, and building alterations are now in progress.

Newcastle has trunk routes to 37 other places, and to mechanize this service means connecting over 1 000 trunk circuits to switch levels, alterations to dialling in the local area, and making '100' available for the operator service in lieu of '0' as early as possible, thus releasing '0' for Grace. The introduction of trunk mechanization and subscriber trunk dialling is expected to save about 12% of the full-time operators, and a further substantial saving

will occur when the service is extended to the satellite exchanges and adjacent towns of Durham, North Shields, etc.

The benefits of automatic over manual signalling seem to be as follows:

(a) Increased speed of establishing calls.
(b) It is very important for follow-on calls, and entails the more or less simultaneous release of the whole train of switches on clearing.

Thus under subscriber trunk dialling, the public should have a quicker service, and the Post Office should be able to make substantial economies in full-time operating staff, usually on a 24-hour basis, and in welfare accommodation. With the advantage so overwhelmingly in favour of subscriber trunk dialling, is there any reason why the scheme should have not been introduced earlier? I believe that the trunk traffic is only a small percentage of the total traffic, but quicker signalling and charging on a time basis, starting with a few seconds, should do much to encourage the use of the trunk service.

If we take the figure '5' as the mean number on the dial, half a second will elapse during pulsing, and probably half a second on moving the dial from the finger stop to the figure '5'. Thus an average of, say, 1 sec per digit, and, without pause, 10 sec per ten digits, although this is likely to be nearer 15 sec in practice. What is likely to be the average time from the end of dialling to the receipt of ringing tone? In view of the exceptionally fast operating properties of electronic switches, is any development taking place towards faster dialling?

An average trunk call would seem to involve about ten pairs of relay or wiper contacts for the transmission path. Has this given rise to any special maintenance problems?

Mr. S. D. Mellor (*at Manchester*): Subscriber trunk dialling is the final phase of mechanization of the trunk service, which began in this country 20 years ago. The provision of a semi-automatic trunk service with single-operator control (known as 'trunk mechanization') is a necessary preliminary to the introduction of subscriber trunk dialling. The work has been completed at some trunk centres and is in progress in others. It is perhaps not generally realized that it represents a series of engineering operations of very considerable magnitude on the part of both the Post Office and the contractors concerned.

The decision, with which I believe the author was closely associated, to commence the first s.t.d. installations before completion of the trunk-mechanization programme seems to have been a remarkably bold one, since it must presumably have been taken before the earliest of the mechanized trunk units had been brought into service. It has had two effects:

(a) The selection of the early s.t.d. centres, e.g. Bristol, necessarily had to be limited to those where automatic trunk units had been installed.
(b) The number of distant exchanges to which subscribers on the early s.t.d. exchanges can dial directly will increase as further automatic trunk units are opened.

Two other major factors which could affect the speed of introduction of subscriber trunk dialling generally are the rate at which development work can be completed and the extent to which the necessary capital can be made available. Could the author comment on the relative importance of these factors?

In Section 1.2 the figure of 321 million inland trunk calls for 1957 is quoted. It is interesting to compare this with the figures for 1937 and 1947, namely 100 and 205 million, respectively. This represents an increase of over three times in 20 years. The practical as well as the economic difficulties of handling such a rapidly-growing trunk traffic on a manual basis must have provided a powerful incentive to mechanize the trunk service and, in particular, to proceed with subscriber trunk dialling as rapidly as possible. The introduction of subscriber trunk dialling seems likely to lead to still higher rates of growth of trunk traffic in the future because of the stimulus offered by the lower charge rates and the greater ease with which trunk calls can be made.

In Section 3 of the paper, stress is laid on the importance of a national numbering plan, which gives each subscriber a unique national number. Was this principle adopted from the outset in those countries where limited schemes of long-distance subscriber dialling were introduced in advance of a general national scheme? If not, presumably some changes in dialling codes will be necessary as subscriber trunk dialling is extended on a national and international basis. In this country subscriber trunk dialling is being introduced from the outset in accordance with the national numbering plan described in the paper, and this difficulty should not arise. There are, however, a large number of existing short-range dialling codes in existence, which will have to be integrated with the national numbering plan. The problem is particularly acute in this part of the country owing to the existence of a large amount of code-dialled unit-fee traffic between the Manchester director area and the surrounding charge groups such as Rochdale, Bolton, Warrington, Macclesfield, etc., with which there is a large community of interest.

Can the author state whether the large number of translations (1 500) which the magnetic drum type of translator is believed to offer is likely to be a major factor in deciding the type of register-translator to be used? It seems to open up the possibility of materially reducing the scale of provision of intermediate and terminal register-translators.

Mr. N. H. Robinson (*at Manchester*): The method of charging which has been adopted will encourage subscribers to make trunk calls of very short duration. These will not be particularly profitable to the administration because the period required to set up the call and to obtain an answer from the called party will probably on the average be about one minute, but the charge may be only that appropriate to a call of 12 sec duration.

There is a big difference between the charges for calls at the local and those at the lowest trunk rate, i.e. 2d. for 3 min against 2d. for 30 sec in the full-rate period and 2d. for 6 min against 2d. for 45 sec in the cheap-rate period. Is there any reason for this big difference? Subsequent steps between charges at the trunk rates are much smaller.

Is it expected that the introduction of subscriber trunk dialling and pay-on-answer call boxes will increase the difficulties with, or the number of, annoyance calls?

Mr. J. Duff (*at Manchester*): With operator trunk dialling, when congestion occurs on a route due, for example, to a cable breakdown, the traffic (or some of it) can be directed to alternative routes. The author does not state whether this facility will be available with subscriber trunk dialling. It will be rather serious if it is not available.

The subscriber has to dial anything from eight to ten digits. In a recent investigation into the ability of telephonists to remember numbers, it was found that the correctness of remembering numbers of eight, nine or ten digits was far from good. The use of letters as well as figures, except where they were part of the name of the locality concerned, did not help much. One would expect that wrong-number trouble would be high with subscriber trunk dialling, and I should be interested to know whether this has happened at Bristol.

It is apparent that the holding and tracing of a faulty connection will not usually be practicable under subscriber trunk dialling, and that to ensure a good and, as far as possible, fault-free trunk service we shall have to rely on frequent and systematic routine testing. This will have to be supported by sampling of the service on a sound statistical basis. It is quite likely that the number of faults reported with subscriber trunk

dialling will be fewer than under operator dialling because the subscriber, on failing to get his connection, will probably try several times before reporting a fault, whereas the operator will report a fault after a second failure.

The 4 sec release feature associated with the variable number of dialled digits might be productive of trouble. One can easily imagine a subscriber pausing for longer than this during dialling because he has forgotten the complete number. I assume that the problem has not proved to be serious in practice.

Mr. E. P. G. Wright (*communicated*): The author refers to the progressive elimination of the incoming and outgoing operators on trunk connections, but, in many cases, a trunk call will involve a private automatic branch exchange at one or both extremities, and the question must arise whether p.a.b.x. operators are a help or a hindrance in an s.t.d. network. It is evident that these operators cannot be removed without certain technical rearrangements, which would need serious study even though they might not be particularly complex. A decision to allow dialling through to extensions would have an influence on the overall speed of service and on the controversy whether a uniform or an economical numbering scheme should be adopted.

It has been mentioned in the discussion that the introduction of subscriber trunk dialling will involve a reassignment of operators, and that, for a period of years, there should be a surplus, but, in the course of time, there may be such a decrease in the quantity and quality available for employment at p.a.b.x.'s that the demand for p.a.b.x. extension trunk dialling will increase.

Although it seems probable that modifications to allow extension dialling would apply to the local network rather than the toll network, it would be of interest if the author could tell us whether the s.t.d. service is designed to permit extension trunk dialling in the future.

Mr. D. A. Barron (*in reply*): I am grateful to Sir Gordon Radley for the valuable information which he has given concerning some of the dominant administrative considerations governing the formulation and implementation of the s.t.d. scheme, and to Mr. Ray for his concise description of our approach to trunk automation and for the information about the way the scheme is working at Bristol.

Several speakers (Messrs. Morrell, Riley and Twycross) have referred to the question of setting-up times, and especially to the length of interval between the end of dialling and the receipt of tone. I would like to comment first on this, because clearly the achievement of fast trunk switching and signalling on s.t.d. calls is a matter of prime importance. Here, as in all countries, the various types of local switching equipment in use must, for the time being, set limits to the overall speed at which trunk calls can be established. In my view, however, it is imperative that the mechanized trunk system now being superimposed to interconnect the local systems shall be as fast as possible, so that when, in due course, the local systems—e.g. by the introduction of electronic exchanges—themselves become extremely rapid, we shall have a high-speed system overall. Reference is made in the paper to the measures proposed to ensure fast signalling and switching on transit calls. In addition, intensive work is in hand to speed up still further the setting-up times for calls between directly connected areas.

Messrs. Morrell and Duff both mention the 4 sec delay period, but this applies to the release of the register, and does not necessarily involve any increase in setting-up time. Furthermore, this point does not in any case arise for calls to the large towns served by director areas, or to any number comprising 10 digits. It would not be possible to adopt the method Mr. King has mentioned, as the number of digits in a national number can vary even within one particular group. The advantage of having such variability is that the most economical arrangement of local equipment can be adopted initially, modified later as necessary to meet growth of the system.

With regard to Mr. Morrell's point about the range of control of the originating register, 'own-exchange-only' control not only reduces the complexity of the register-translator but has the advantage that changes in routing at any point in the network do not require changes to translation data on a country-wide basis, but only at the register centres immediately concerned with the routing in question. This is of particular importance with respect to temporary routing changes made under breakdown conditions. The adoption of 'own-exchange-only' control does not preclude the use of automatic alternative routing, which will, in fact, be provided at all trunk switching centres. It is factors of this nature which govern the desirable range of control of the originating register-translator and the scale of provision of intermediate and terminal register-translators. This also covers Mr. Mellor's point about the translation facilities available with magnetic-drum type equipment.

With regard to Mr. Crutch's comment about the new coin box, it should be remembered that the fact that a coin-box user cannot hear the called subscriber until a coin has been inserted in the box merely brings the coin-box user into line with an ordinary subscriber, who is debited with the initial call charge immediately the called subscriber answers. With periodic metering the initial charge is limited to one unit fee regardless of distance. Thus the sum involved is 2d. for an ordinary subscriber and 3d. for a coin-box user even on a long-distance trunk call.

In reply to Mr. Tolley, the total capital cost of the equipment needed to provide subscriber trunk dialling for 75% of all trunk calls by 1970 has been estimated to be about £35 million. By 1970 the trunk traffic will probably have grown to nearly double its present volume, and owing to the reduction in the number of operators, the saving has been estimated to be about £15 million a year. Subscriber trunk dialling therefore shows substantial overall savings which are being passed on to the subscriber in lower call charges.

It seems unlikely that the reduced standard of provision of switches adopted for local exchanges will be applied in automatic trunk exchanges. The whole question of the grade of service on the trunk network will need to be reviewed in the light of s.t.d. experience and to take account of the plans for a fast trunk switching network with automatic alternative routing.

Mr. Smart asks about the use of cold-cathode tubes for counting and storing in the Bristol register-translators. Apart from the maintenance advantages he mentions, we were considerably influenced by the fact that such techniques were the only ones of which we had obtained several years' experience in experimental installations carrying live traffic. The decision to use these techniques at Bristol has been amply vindicated by the excellent performance obtained, but the later designs of register-translator for use in director areas will be based on the magnetic drum. Other technical solutions to the register-translator problem, including that mentioned by Mr. Pollard, will, no doubt, become available in the future, and will be considered on their merits.

With regard to Mr. Roebuck's query, the position is that metering ceases either when the calling subscriber clears the connection, or after a delay of a few minutes if the called subscriber clears while the calling subscriber is still holding the connection. Short-duration interruptions, e.g. owing to the called subscriber temporarily replacing the telephone and later removing it again to continue the conversation, do not affect metering, which continues throughout.

Mr. Glanfield has made some interesting points about figure

versus mixed letter and figure codes. It was decided, after full consideration of all the relevant factors, that the national interest would be best served by using mixed codes in the way described. Subscribers at Bristol seem to have found the numbering scheme simple to understand, and even in the first week were dialling 95% of their calls to exchanges accessible by subscriber trunk dialling. The use of letters is, of course, essential for a simple dialling procedure for s.t.d. calls to director exchanges, and in reply to a recent inquiry 82% of Bristol subscribers say they find the letters helpful. The numbering scheme we have adopted is not expected to cause any serious difficulties in subscriber dialling from the Continent, although admittedly some letter-figure cross reference will be necessary for the foreign subscriber using an all-figure dial.

In the investigation mentioned by Mr. Duff, the telephonists had to carry 8 to 10 digits in their heads, but subscribers usually have the dialling code and number before them as they dial. Service observations at Bristol show that subscribers make dialling errors on about 5% of s.t.d. calls, but most of these lead to the number-unobtainable tone. Less than 1% of s.t.d. calls get wrong numbers due to dialling errors.

Mr. Wright raises the question of direct dialling to p.a.b.x. extensions, and this clearly could be a very useful facility. However, it would mean either encroaching on exchange numbering capacity or making modifications to directors and register-translators to cope with additional digits. These consideration will probably limit the extent to which the facility can be give The problem is currently being studied.

Mr. Mellor's query about the rate of introduction of the ne scheme is largely covered by Sir Gordon Radley's comment The availability of accommodation will, in some cases, be further limiting factor.

In reply to Mr. Robinson, experience at Bristol does no suggest that the number of very short calls will be sufficient t upset the economics of subscriber trunk dialling. There ha been a reduction in the average duration of trunk calls, but onl by 16%, whereas the number of calls has increased by 40%

The extensive area available at 2d. for 3 min results from th principle of allowing calls to adjacent groups at the local rat This is the simplest way of overcoming the cross-boundary ca problem which arises with a system of group charging for shor distance calls. The low rate is achieved by averaging costs fo the whole of the local call area. With present techniques fo providing trunk circuits the cost rises comparatively slowly wit distance in the high-frequency range.

There is no reason to expect that subscriber trunk dialling wi increase the number of annoyance calls. With the new pay-o answer coin box the caller cannot hear the called person's voic without inserting coins, and this feature may, in fact, reduc annoyance from call offices.

DISCUSSION ON 'THE CAPACITANCE BETWEEN ELECTRODES IN THE PRESENCE OF SPACE CHARGES'*

Mr. Robert Willett (*Germany: communicated*): I can follow Dr. Bull's argument up to and including Section 3. The result of Section 4, $C_C = 0$ in the space charge regime, rests on the assumption that the cathode current is not dependent on V but is constant. This assumption is acceptable for the saturated regime (also, in what follows for the conditions assumed by Bull). In the space-charge-limited regime the space charge density increases by a negative charge on applying a slight increase in voltage δV to the anode [see Barkhausen, 'Elektronenröhren' (Hirzel, Leipzig, 1953), vol. 1, p. 56]. Since the electron speed also increases everywhere, the addition to the total negative space charge must be brought about by the loss of negative charge from the cathode. The cathode is in a position to do this, since in the space-charge-limited regime it has a reserve of emitting capacity. This situation must be taken into account by means of a correction factor of the form $\partial Q_S/\partial V$ in eqn. (3). Since also in the space-charge-limited regime $(\partial V/\partial x)_C \equiv 0$, it can be concluded that C_C is not zero.

Dr. C. S. Bull (*in reply*): In Section 4, it has not in fact been assumed, even tacitly, that the anode current is independent of anode voltage. To justify this remark, using Child's approximation, we can proceed as follows:

Let the potential V be increased to $V + \delta V$ at a small arbitrary rate $\partial V/\partial t$ so slowly that at all times electron inertia effects are negligible.

At whatever value of δV this process may be arrested, we shall find that:

(*a*) The current from anode to cathode, and of course in the anode and cathode leads, is calculable from Child's equation.

(*b*) The convection current of electrons from the cathode is equal to the conduction current in the cathode lead, and $(\partial V/\partial x)_C \equiv 0$.

(*c*) The convection current of electrons to the anode is equ to the current in the anode or cathode leads, and also the electr field at the anode, $(\partial V/\partial x)_A$, has increased to a new value, i contrast with the field at the cathode, which has remaine constant at zero.

From (*a*) and (*b*) we conclude that at all times the convectio current leaving the cathode is exactly equal and opposite to th conduction current in the cathode lead, and therefore that n net charge leaves the cathode, even though the current do change with V. The cathode is therefore not making any co tribution to $\partial Q_S/\partial V$, so that eqn. (3) is valid for space-charg limited conditions.

From (*a*) and (*c*), however, it is seen that a displaceme current $\epsilon_0 \frac{\partial}{\partial t}(\partial V/\partial x)_A$ exists at the anode during the time th $\partial V/\partial t \neq 0$. This displacement current indicates that a positi charge is collecting on the anode.

Since, in order to define or measure the capacitance betwe the two electrodes, namely cathode and anode, equal curren must flow in both leads at all times, this positive charge accum lates at the anode on account of a failure of electrons to rea it by convection. These electrons, leaving the cathode a failing to reach the anode while $\partial V/\partial t \neq 0$, build up the increas space charge and account for the term $\partial Q_S/\partial V$ in eqns. (and (10).

The calculation of Section 4 then shows that $C_C = C_A =$ even though the current depends on V.

Similar conclusions would be valid if electron emissi velocities were negligible but not zero. If, however, the emissi velocities are appreciable or control the current, as at low ano voltages or in the retarding region, quite different results cou be expected from those given in the paper.

* BULL, C. S.: Paper No. 2333 R, July, 1957 (see **104** B, p. 374).

621.395.61 : 621.395.665

The Institution of Electrical Engineers
Paper No. 2867 E
Mar. 1959

DESIGN OF AN AUTOMATIC SENSITIVITY CONTROL FOR A NEW SUBSCRIBER'S TELEPHONE SET

The British Post Office 700-Type Telephone

By F. E. WILLIAMS, M.Sc.(Eng.), and F. A. WILSON, Associate Members.

(*The paper was first received* 10*th October, and in revised form* 29*th December*, 1958. *It was published separately in March*, 1959.)

SUMMARY

The aim of the telephone designer has generally been the attainment of greater sending and receiving sensitivity, so that conversations may be carried on over connections of greater loss. With the telephone sets hitherto available, the sensitivity has not been so high as to give excessively loud reception on short connections. However, recent developments, particularly in the electromagnetic design of telephone receivers, have made available substantial increases in sensitivity, and the new British Post Office subscriber's telephone set (the 700 type) is more sensitive on long lines than its predecessor (the 300 type) by some 4 dB in both sending and receiving directions. This enhanced sensitivity, which is of great advantage in permitting the use of longer or smaller-gauge subscribers' local lines without degradation of performance, is liable if uncontrolled to be an embarrassment on short lines, and, in fact, subjective tests carried out in the field and in the laboratory have established that, on short connections, most subscribers would find this sensitivity too high.

The problem is discussed, and it is shown that the best solution is the incorporation in the telephone set of an automatic regulating device which reduces the sending and receiving efficiencies and the sidetone level when the line current is high, while leaving the full sensitivity unimpaired on longer lines. The regulator makes use of the non-linear voltage/current characteristics of selenium-rectifier elements, biased in the forward direction by a control voltage derived from the voltage drop across a metal-filament lamp in series with the line. These elements are used to provide shunt losses on short lines of up to 6 dB on sending and 4 dB on receiving, and also to reduce sidetone.

Details are given of the performance of the new Post Office Telephone No. 706 which embodies such a sensitivity control as an integral part of the telephone set. It has as good a performance on a 1-kilohm local line as the 300-type set had on a 660-ohm line; at the other extreme, on very short lines, the sensitivity of the new set is, owing to the action of the regulator, no greater than that of the old set.

(1) INTRODUCTION

Unlike the majority of sound-reproducing devices in everyday use, the subscriber's telephone set does not include any form of electrical amplification whose gain can be regulated to provide the desired output sound level. In the evolution of the telephone system, the designers of sets have striven to obtain the greatest possible sending and receiving sensitivity, and the planners of the telephone system have ensured that the transmission loss between any two subscribers in the network shall always be within limits such that a reasonable minimum standard of communication is maintained.

In the British Post Office telephone network this minimum standard corresponds in sensitivity (assessed by loudness balancing against a high-quality reference system) to 6 dB loss introduced into a one-metre air path; in this context a one-metre air path implies monaural listening to a talker at a distance of one metre in echo-free surroundings. The evolution of this transmission standard has been discussed elsewhere in some detail.[1] It will suffice to say it is defined physically by a telephone connection, each end of which consists of a Telephone No. 162 connected by 2·56 miles of local line of 10 lb cable (of 450-ohm loop resistance) to a 50-volt non-ballast exchange bridge with a 27 dB attenuator, representing losses in junctions and trunks, including switching and mismatch losses, inserted between the two bridges. Other combinations of telephone set, local line and feeding bridge have been rated to find the limiting local lines which give the same performance as the standard. For example, with the type-300 telephone the permitted line limit is 660 ohms of $6\frac{1}{2}$ lb cable for 50-volt non-ballast exchanges.

When two subscribers, each having very short lines to the same exchange, are connected together, the overall sensitivity of the connection will be, with type-300 telephone sets, nearly 40 dB more than the minimum standard. The possible range of sensitivity is illustrated in Fig. 1, together with some recently determined data on subjects' sensitivity preferences. These data, which are based on experiments carried out at the Post Office Engineering Research Station, show quite good agreement with earlier work in Germany.[2] With the type-300 set the maximum desirable sensitivity of about 31 dB relative to a metre air path is already exceeded by some 2 dB, and any further increase in sensitivity is liable to be an embarrassment on short connections. The need exists for a telephone set of high efficiency, however, to enable longer (or lighter-gauge) subscribers' lines to be used without degradation of performance, and the target set by the requirements of the network is for a local line limit equivalent to 1 kilohm of $6\frac{1}{2}$ lb cable. With telephone sets of this enhanced sensitivity, the received speech level on a short connection may be as much as 10 dB greater than that generally acceptable.

The following is an outline of the problems encountered in the development of the new telephone set:

(*a*) Is it practicable, by the application of modern materials and techniques, so to improve the performance of the telephone set for both sending and receiving* that the local line limit can be raised to 1 kilohm of $6\frac{1}{2}$ lb cable?

(*b*) With telephones of this increased sensitivity will subscribers find reception and sidetone unacceptably loud on short lines?

(*c*) What features can be incorporated in the telephone to reduce the sensitivity and sidetone on short lines, while leaving the full sensitivity unimpaired on long lines?

The idea of providing within the telephone set a sensitivity control dependent upon line current is not new, and telephones incorporating a measure of control have been described elsewhere.[3,4] However, it will be seen that the excess sensitivity on

* The principle of separately rating sending and receiving does not allow advantage to be taken (e.g. by raising the line limit) of better receiving performance if the limit is already set by sending. Recent improvements in the electromagnetic design of telephone receivers have led to a large increase in sensitivity, which has not been matched by any corresponding increase in sensitivity of the carbon microphone, and it could be argued that the standard should be changed to allow worse sending but better receiving efficiency. The consequence of such a change of standard would be that, on calls between a new and an old telephone, the transmission standard for the direction of speaking from new to old would be degraded. This degradation would persist, though for a decreasing number of calls, throughout the period taken to change from old to new telephones throughout the network. Furthermore, eventually speech levels on lines would be lower and speech/line-noise ratios less favourable.

Mr. Williams and Mr. Wilson are at the Post Office Engineering Research Station.

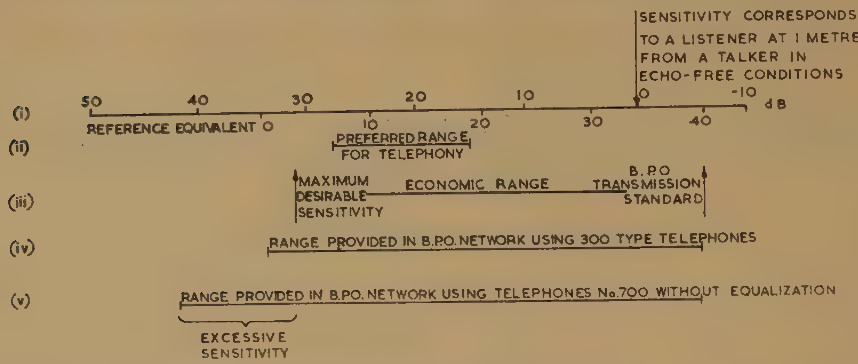


Fig. 1.—Telephone sensitivity ranges.

The 'reference equivalent' is the rating on a loudness basis relative to the C.C.I.T.T. Reference System (S.F.E.R.T.).

short lines of a telephone which meets requirement (*a*) is so high as to demand a new approach to the problem of sensitivity control.

In the paper are discussed the evolution of an experimental new British Post Office telephone set (the Telephone No. 700)[5] of higher efficiency; the determination by subjective methods of the highest level of sensitivity acceptable to the user; the development of a regulator for automatically reducing the sensitivity of the new set when it is used on short lines; and, finally, the circuit design of a new set (the Telephone No. 706) embodying the regulator as an integral part of the set circuit.

The scope of the paper is restricted to transmission performance, and does not include any discussion of signalling or the physical layout of the components in the telephone set.

(2) THE EXPERIMENTAL TELEPHONE No. 700

In the evolution of the new telephone set the immediate target was to raise the local line limit to 1 kilohm of $6\frac{1}{2}$ lb cable. This required an increase in sensitivity over the existing 300-type set of some 4 dB both on sending and receiving. Under the auspices of the British Telephone Technical Development Committee[6] (the joint organization of the British telephone industry and the Post Office), the development of new capsule-type receivers and transmitters was commenced. The receiver development proved the more fruitful, and by 1952 a British manufacturer had successfully produced a new receiver of rocking-armature design[7] employing modern magnetic materials and embodying acoustic equalization. By separating the acoustic and electromagnetic functions a considerable increase in sensitivity was obtained, together with an improved frequency response. Fig. 2 shows its sensitivity relative to that of the receivers type 1L and 2P used in the early and later 300-type telephones. This new receiver has now had several years of rigorous testing under service conditions.

Transmitter development proved slower and more beset with difficulties, and it became evident that, while it would be practicable to produce an acoustically-equalized transmitter with a smoother frequency response and less amplitude distortion than the existing immersed-electrode transmitter inset No. 13, no substantial gain in sensitivity was likely to materialize. A carbon transmitter of new design must of necessity undergo prolonged field trial to ensure that it is free from any tendencies to become noisy, i.e. to 'fry', or to 'pack', or to increase its resistance with age due to burning of the granules and formation of carbon dust. It was decided to proceed with the production of the new telephone without waiting for the completion of the trials of a new transmitter, and the design of the new telephone handset is such as to allow for the introduction of a new design of transmitter at a later stage.

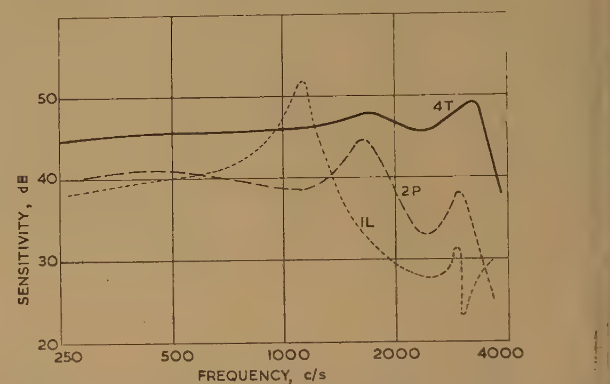


Fig. 2.—Sensitivity of British Post Office receiver insets 1L, 2P and 4T in decibels relative to 1 dyne/cm² per $\sqrt{}$(mW) available power.

With the conventional anti-sidetone induction coil under balanced conditions, only part of the power available from the transmitter is sent to the line, the remainder being dissipated in the balance.[8] The ratio

$$\frac{\text{A.C. power into line}}{\text{A.C. power into balance}}$$

from the transmitter when sending is also equal to the ratio

$$\frac{\text{A.C. power into transmitter}}{\text{A.C. power into receiver}}$$

from the line when receiving, and will be termed the *Y*-ratio of the circuit. For maximum overall efficiency (sending plus receiving) *Y* should be unity, but by making *Y* greater than unity, a moderate increase in sending efficiency can be obtained, at the expense of a greater loss of receiving efficiency. Ample sensitivity is now available from the new receiver, and so that the telephone shall be approximately balanced in its sensitivity improvements relative to the 300-type for both sending and receiving, the *Y*-ratio can be raised to about 3. This results in an increase in sending efficiency of 1·8 dB and a decrease in receiving efficiency of 3 dB.

The transmitter inset No. 13 gives its highest efficiency when held so that the front face is inclined at about 40° to the vertical. The existing Post Office telephone handset design is such that, when held normally, the face of the transmitter is inclined at about 20° to the vertical. The new handset, in addition to embodying a new mouthpiece of more open design than the horn type, is so shaped that the transmitter is inclined at the more favourable angle; the combined effect of this change of angle, increase of *Y*-ratio, and adoption of a closed-core induction coil

of greater efficiency, is to give the required increase in sending efficiency.

Fig. 3 illustrates the performance of the experimental Telephone No. 700 on a loudness basis relative to that of the 300-type set,

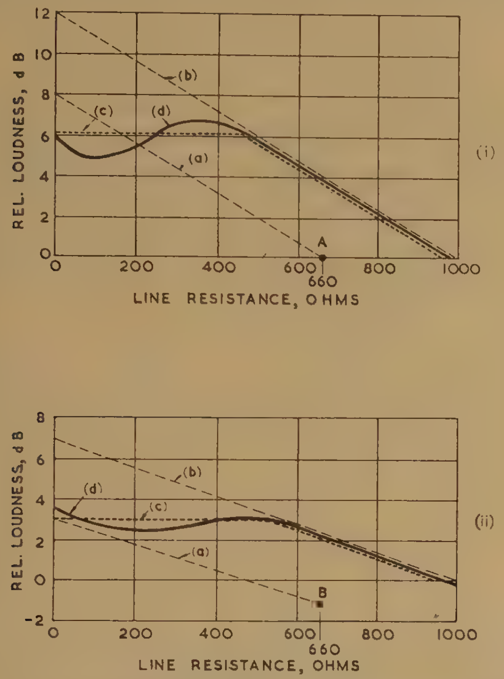


Fig. 3.—Relative loudness of 300-type and 700-type telephone circuits.

The telephone set is connected by local line of $6\frac{1}{2}$ lb/mile cable, via a 50-volt Stone bridge with 200 + 200-ohm relays, to a 600-ohm junction. The curves show the effect of line length on loudness for (*a*) 300-type telephone, (*b*) experimental 700-type telephone, (*c*) best economic characteristic for 700-type telephone, (*d*) telephone No. 706.
A corresponds to reference equivalent 12 dBW, AEN 19 dB.
B corresponds to reference equivalent 1 dBW, AEN 5 dB.
(i) Sending.
(ii) Receiving.

and shows that the primary design target has been met. On a local line of 1 kilohm of $6\frac{1}{2}$ lb cable the loudness efficacy on sending is the same as that of the 300-type set on a 660-ohm line, and on receiving it is 2 dB better. (Fig. 3 also shows the performance of the Telephone No. 706 with regulator; this will be referred to later.)

With two of these telephones connected by a very short line, the overall sensitivity of the connection will be 8 dB higher than for two 300-type sets. Fig. 1 shows that the maximum desirable sensitivity figure of 31 dB relative to a one-metre air path is now exceeded by some 10 dB. The next stage in the development of the new set was to ascertain subjectively whether the sensitivity level was, in fact, excessive in practice.

(3) NEED FOR REGULATION

The reaction of users with short lines to the increased sensitivity of the experimental Telephone No. 700 was studied in two ways: first, by a field trial conducted at two Telephone Manager's Offices, and secondly, by an 'opinions' test conducted in the laboratory using untrained subjects drawn generally from the staff at the Post Office Engineering Research Station.

(3.1) The Field Trial

Two Telephone Manager's Offices were selected for the trial, one with a parallel-feed manual switchboard (type CB9) and the other with a 50-volt automatic exchange (type U.A.X.13). The existing 300-type telephones throughout the offices were replaced by the new type, and, after a period of one month, the users were interviewed singly and asked whether they had noticed any differences in the new telephone compared with the old. The total number of sets involved in the offices was about 400, and the staff concerned were partly clerical and partly engineering; they were not given any lead as to the differences to look for. The minimum of prompting was used, and the subjects were never in any way pressed to give an opinion on any point on which they appeared to have no strong feeling.

On internal calls in these offices the telephones were working under conditions of maximum sensitivity, since none of the lines exceeded 25 ohms in resistance. With regard to room noise, however, these offices were probably quieter than the average owing to the location of the buildings and the fact that, since the weather was cold, all the windows were kept shut.

Analysis of the opinions showed very clearly that the users found the sensitivity excessive; some 75% of the subjects said they found the telephone too loud on internal calls, and 49% of the subjects found it too loud on external calls. Some of the staff had actually taken steps to reduce the loudness, either by inserting a resistor at the terminal block in series with the line or by placing several thicknesses of paper under the ear cap. A common attitude was that the subject was prepared to accept the excessive loudness on local calls for the sake of the improved reception, particularly on calls which had previously been difficult. The better speech quality was commented on, and the majority expressed a clear preference for the new set against the old. A fair proportion of the subjects who had at first disliked the high sensitivity had found later that they were getting used to it and regretted not being able to have the new telephone in their own homes. It should also be noted that some of the subjects who found the level satisfactory were rather hard of hearing.

Sidetone was not commented on critically, but it seems probable that many of the subjects may not have fully understood the difference between this and receiving loudness.

Measurements made at one of the Telephone Manager's Offices of the outgoing speech voltages before and after the installation of the new sets showed an increase in the average from $-7\cdot3$ to $-6\cdot1$ dB relative to 1 volt. The fact that the whole increase of 4 dB to be expected from Fig. 3(*a*) is not realized may be due to several factors, the chief of which is probably the effect of the loud sidetone in reducing the talker's vocal level.

(3.2) The Laboratory 'Opinions' Test

To assess the performance of the experimental telephone when used on short local lines under more controlled conditions than those which existed during the field trials, a laboratory experiment was arranged in which 32 untrained subjects drawn generally from the staff at the Post Office Engineering Research Station were required to carry out a conversational test,[9] involving completion of a series of tasks, over the circuit under test. At the completion of each conversation the users were asked to indicate in which of the following five categories they placed the call:

	Opinions score
Excellent	4
Good	3
Fair	2
Poor	1
Bad	0

Fig. 4 shows mean opinions score plotted against overall loudness efficacy, which is the result of a loudness-balance comparison conducted under fixed talking conditions (vocal level and lip position) by a trained crew against a high-quality

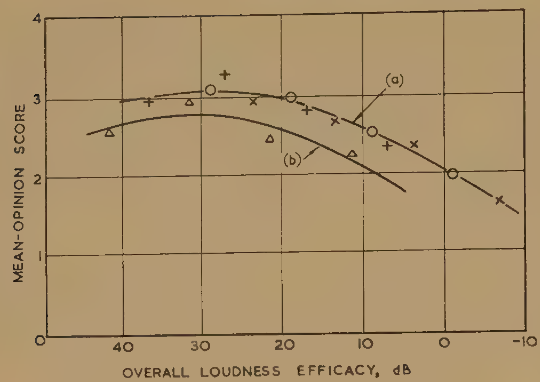


Fig. 4.—The experimental Telephone No. 700.

The mean opinion score is given as a function of overall loudness efficacy, in decibels relative to a metre air path, for subscribers' lines of 0, 270, 660 and 1 000 ohms $6\frac{1}{2}$ lb cable. 50 dB room noise (Hoth spectrum) is present at each end.

△ Zero line. ○ 660-ohm line.
+ 270-ohm line. × 1-kilohm line.

reference channel; the reference condition corresponds in sensitivity to a one-metre air path between talker and listener, and it is again evident that sensitivities approaching 40 dB relative to a metre air path are excessive.

This form of presentation permits certain general conclusions to be drawn. The results for lines of 270, 660 and 1 000 ohms of $6\frac{1}{2}$ lb cable lie closely about a single curve [curve (*a*)], and it may be inferred that the opinion is influenced only by the overall loudness efficacy, and that the sidetone sensitivity which corresponds to subscribers' lines over this range has no practical influence on the opinion. The sidetone sensitivities (expressed as loudness efficacies relative to metre air-path reference conditions) for these three conditions do not exceed +25 dB. Curve (*b*) shows the results for zero-loss subscribers' lines, and it is clear that the opinion scores are lower (for any given overall sensitivity); this may be attributed to the greater sidetone sensitivity of +33 dB.

It appears that a sidetone sensitivity of up to +25 dB will not affect the users' opinions, but that one of +33 dB will have a marked adverse effect. This is in agreement with the result of earlier work on the effects of sidetone carried out at the Post Office Engineering Research Station, in which a sidetone sensitivity of +25 dB was unnoticed by as many as 30% of the subjects and objected to by only 3%, whereas a sidetone sensitivity of +33 dB was unnoticed by only 20% and objected to by as many as 18%.

The Telephone No. 700 circuit was designed for the best electrical sidetone attenuation possible on the longer lines using a conventional telephone transmission circuit. A study of Fig. 5 reveals that this necessitates somewhat higher sidetone levels on the short lines. This Figure shows, in full line, the curves of impedance expressed as R and X 'seen' from the line terminals of a subscriber's installation towards the exchange, for local calls, (*d*), and for several classes of junction circuit, (*a*), (*b*) and (*c*), outgoing from the exchange with typical subscribers' cables of $6\frac{1}{2}$ lb/mile conductor weight, varying from 0 to 1 kilohm resistance (0–3·7 miles). The values shown are for one type of exchange transmission bridge (Stone) and at one frequency only, and are indicative of the large range of impedances to be catered for in sidetone balancing. The constant-sidetone circles show the sidetone attenuation of an experimental 700-type circuit for a balance point at $510 - j430$ ohms. A change of balance circuit so that the point occurred at, say, P_1 would result in an electrical sidetone performance more constant with line length, but would give a worse performance on the longer lines. One way of avoiding this dilemma is to balance as shown and provide an additional means of sidetone attenuation on the shorter lines. It will be shown later that this can be accomplished by a sensitivity regulator.

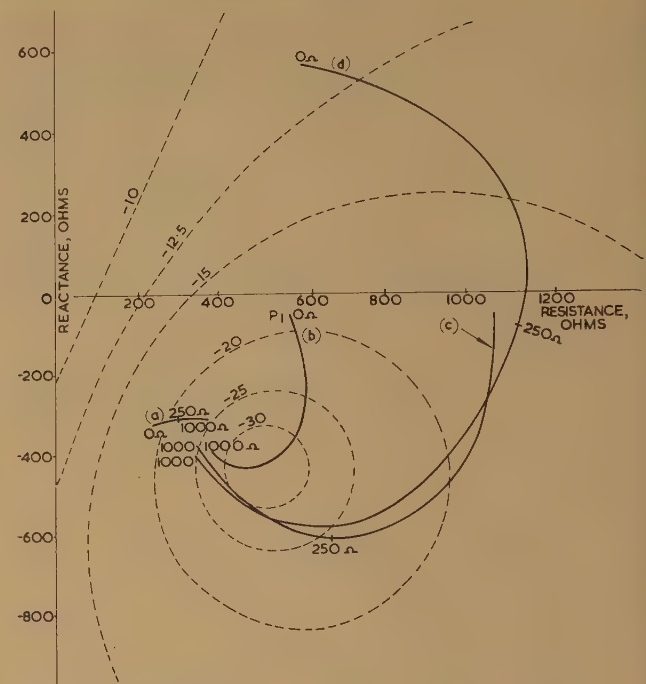


Fig. 5.—Impedance at line terminals of a subscriber's set (at 2 000 c/s).

With a Stone-type transmission bridge and subscriber's line of $6\frac{1}{2}$ lb/mile cables connected to (*a*) unloaded junction, (*b*) amplified junction, (*c*) loaded junction, and (*d*) local call to 300-type telephone on zero line.

The figures on circles represent sidetone attenuation in decibels relative to the potential difference across a 200-ohm receiver when the transmitter e.m.f. is 1 volt.

(3.3) Summary of Results of Trials

It may be inferred from the results that the overall sensitivity of the experimental telephone set on zero-loss subscribers' lines and zero-loss junction exceeds the optimum sensitivity by some 10 dB, and that, under these conditions, the sidetone is excessive by at least 8 dB. The user trials show that, although many would be prepared to tolerate these high sensitivities for the sake of their advantage under unfavourable line conditions, it is nevertheless desirable that the new telephone should embody some device for reducing its sensitivity on short lines, while retaining the maximum possible sensitivity on long lines.

(4) DESIGN OF THE AUTOMATIC REGULATOR

(4.1) Desired Characteristics

The apportionment of the 10 dB reduction between 'send' and 'receive' is a matter for some compromise. An even division is desirable to avoid unbalance between the two directions of transmission when there is an older type of set at the distant end. However, excessive sensitivity is more of an embarrassment on sending than receiving, because of its effect on other parts of the system, e.g. in crosstalk and in overloading the carrier-system amplifiers, and it has already been shown that the sensitivity of the present type-300 set is some 2 dB too high on sending. The best choice appears to lie in 6 dB reduction of sending sensitivity and 4 dB reduction of receiving sensitivity. This leaves the sensitivity of the new set on zero-loss line as equal to that of the type-300 set on receiving, and 2 dB less on sending.

The action of the regulator should become progressively less as the line length is increased, becoming negligible at about 500 ohms, as shown by the dotted line (*c*) in Figs. 3(i) and (ii); the results shown are for one gauge of cable (6½ lb per mile) only, but the ideal requirements for other gauges of cable will differ only slightly from these.

(4.2) Methods of Sensitivity Regulation

Some devices are already in use which provide a small measure of control by regulating the line current; an example is the ballast or barretter type of transmission bridge. However, the required range of control is not attainable by such means, and the effect is confined to sending only. It is not a satisfactory solution to provide the regulating means within the exchange transmission bridge because the control will then be applied indiscriminately to existing sets which are not excessively sensitive.

The requirements could be met in part by the inclusion in the telephone set of a fixed-loss pad which could be connected into the circuit by means of a link when the set was installed on short lines; such a device might consist simply of two resistors providing predetermined amounts of sending and receiving loss with reduction of sidetone. This is not a wholly satisfactory solution, although it meets the requirements on very short lines; it entails applying an unnecessarily high loss on lines between, say, 200 and 400 ohms, a range embracing the average of the British network.

Moreover, it is the private branch exchange (p.b.x.) which poses the more difficult problem. The length of the extension lines may be very short, and yet the p.b.x. may have a long line to the main exchange. Extension-to-extension calls on the same p.b.x. may therefore be excessively loud, and sensitivity control will be necessary; but when an extension telephone is connected through to the main exchange the p.b.x. transmission circuits are usually switched out and the telephone is effectively working on a long direct exchange line.

It might appear that the requirements could be met, at least in part, by means of a voice-operated volume control, in which the level of the signal itself determined the loss required, but this is not a satisfactory solution. To achieve constant sending level, such a device would have to be installed at the exchange end of the line, because it is at this point that a constant speech level is required. For receiving, the device would have to be connected within the telephone instrument to maintain a constant loudness from the receiver; sidetone improvement would be gained only while there was a received signal, and its value would be limited to that of the receiving regulator loss.

All these difficulties can be overcome by the use of an automatic sensitivity regulator in the telephone set which adjusts its loss according to the length of the local line. The degree to which the ideal requirements can be approached will naturally depend upon the complexity and cost of the device.

(4.2.1) Requirements of an Automatic Sensitivity Regulator.

The desirable features of the regulator may be summarized as follows:

(*a*) It should introduce losses in the sending, receiving and sidetone paths of the instrument according to the resistance of the local line, as shown in Fig. 3.

(*b*) Harmonic distortion should not be introduced.

(*c*) The performance should be unaffected by variations in other components of the telephone instrument—in particular, by the resistance of the carbon-granule transmitter.

(*d*) The regulator components should be robust and stable, and the design should be such that failure of a component must not put the telephone completely out of action.

(*e*) It must be compact and of low cost.

It is clear from Fig. 3 that the regulator must be a variable-loss device which is capable of sensing the resistance of the local line and adjusting its loss accordingly. An approximate measure of line resistance is provided by the magnitude of the line direct current, which ranges from about 100 mA on very short lines to about 30 mA on the highest-resistance lines. (The word 'approximate' needs some qualification; in the British telephone network some of the older exchanges still in use—mainly manual and p.b.x.—have line-current ranges very different from this.)

(4.2.2) Use of Non-linear Elements.

Some of the possible methods of providing automatic regulation by line-current sensing are eliminated by considerations of cost. For example, line-current saturation of the magnetic core of the receiver to reduce its sensitivity, or of the core of the induction coil to reduce its efficiency, would require specially developed components, and the former would still require some additional method of regulation to be provided for the sending path.

Losses within a telephone set are most easily obtained by introducing series or shunt resistance or capacitance. The element must be non-linear, and the obvious choice is a non-linear resistor which falls in resistance with increase in current or voltage, used as a shunt element. Consider a 100-ohm generator working into a matched 100-ohm load R_L (these values are of the order of those encountered in a telephone set with the carbon transmitter sending into the line via the induction-coil circuit). Let the power in R_L be reduced by means of a shunt R_s. The regulator is required to reduce the sending level on zero-loss line by 6 dB, for which $R_s = 50$ ohms. On long lines the loss must be negligible—say, not in excess of 0·3 dB, for which $R_s = 1{\cdot}5$ kilohms. Thus a line-current fall from 100 to 30 mA must be translated into a shunt-resistance rise of the order of 1 : 30.

Excluding items which might be classed as too expensive for this application, the available types of non-linear element are the thermistor, the silicon-carbide disc, and copper-oxide and selenium rectifiers. The thermistor, which is a thermal device relying upon the electric heating of a bead of material having a negative resistance/temperature characteristic, would appear to be very suitable, since, owing to its slow heating, it does not introduce distortion at audio frequencies and it is capable of resistivity changes in excess of that required. For example, standard production items can give resistivity changes of as much as 400 : 1 for some 60 mW dissipation. It seems very probable that, in the future, thermistors may be developed which will fully meet the requirements, but, for the present application, the thermistor has not been seriously considered because (*a*) special miniature types would have to be developed, (*b*) the spread of characteristics for normal production items is rather large, (*c*) their cost is probably a little high, and (*d*) their inability to withstand surges would necessitate some additional element for surge suppression.

The remaining non-linear elements—silicon carbide, copper oxide and selenium—may now be considered together, for their instantaneous voltage/current relationships follow approximately the same law:

$$I = kV^n$$

where n is the power index of the material.

At any point on the voltage/current curve the d.c. resistance is given by V/I, and the a.c. resistance to small signals by the slope

$$\frac{dV}{dI} = \frac{1}{nkV^{n-1}}$$

As will be seen later, for copper oxide and selenium this resistance is shunted by inherent capacitance.

Let us consider two voltages, V_1 and V_2.

$$\frac{\text{A.C. resistance at } V_1}{\text{A.C. resistance at } V_2} = \left(\frac{V_2}{V_1}\right)^{n-1}$$

If the line-current change from 100 to 30 mA be converted to a voltage change of the same order (e.g. from a small resistor inserted in series with the line), so that

$$\frac{V_2}{V_1} = \frac{100}{30}$$

then, for an a.c. resistance ratio of 30 : 1,

$$\log 30 = (n - 1) \log 3{\cdot}33$$

and $n = 4$ approximately.

A disadvantage of silicon-carbide discs is that a.c. resistances as low as 50 ohms at the direct voltages available are only obtained in elements with comparatively low power indices. Rectifier elements of copper oxide or selenium are better in this respect, and they also give higher power indices; since they are cheap, compact and reliable and able to withstand high surges they are very suitable for this application. They are unidirectional devices, and as the direction of line-current flow cannot be foreseen they must be used in parallel-connected pairs of opposite polarity.

(4.2.2.1) *Derivation of Control Voltage.*

The direct-voltage drop across the transmitter is unsuitable for biasing the shunt elements, because the resistance of the transmitter falls as the current rises, and the direct voltage therefore changes very little; typical resistances for the transmitter inset No. 13 are 150 ohms at 30 mA and 60 ohms at 100 mA, giving a voltage change from 4·5 to 6 volts. The transmitter resistance also tends to increase with ageing and varies considerably from one inset to another.

A suitable control voltage can be obtained from a series resistor inserted in the feeding current path. This current-sensing resistor creates a transmission loss, and to avoid degrading transmission on long lines its resistance must be small. On short lines, however, the voltage drop across it must be sufficient to drive the non-linear elements to the desired point on their voltage/resistance characteristic. These conflicting requirements can be met by the use of a non-linear resistor, the resistance of which rises with current, with the following advantages:

(*a*) The control voltage range is increased.
(*b*) The transmission loss caused by the resistor is least at low line currents, i.e. on long lines.

Suitable resistance/current characteristics for the sensing resistor are obtainable from a tungsten-filament lamp.

(4.2.2.2) *Avoidance of Non-linear Distortion.*

In the circuit shown in Fig. 6(*a*), if a rectifier element is connected across the points A and B and a direct current flows through it in the forward direction, an a.c. shunt path is provided across A and B. Unless the direct current is greatly in excess of the alternating current, distortion of the a.c. waveform in the load is inevitable. For a single-plate rectifier shunting a 600-ohm load as shown, and d.c.-biased to produce 6 dB loss, the second harmonic produced is about 25% of the fundamental at the speech levels encountered on a short subscriber's line. Distortions of this magnitude are appreciably higher than those generated within the carbon transmitter (which are about 14% total harmonics)[10] and would be noticeable, particularly on sidetone.

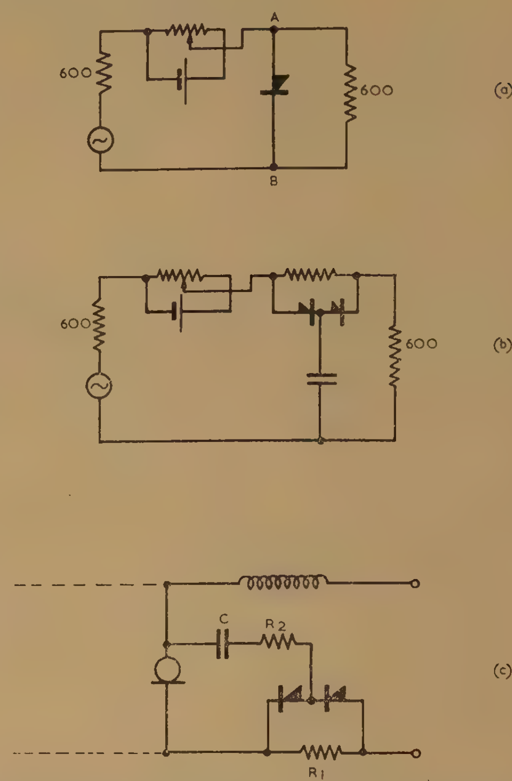


Fig. 6.—Reduction of distortion in shunt rectifier paths.
The circuit arrangement of (*b*) introduces less non-linear distortion than that of (*a*). In (*c*) it is shown as a transmitter shunt.

Distortion can be brought to a much lower value by working the rectifiers in the push-pull arrangement of Fig. 6(*b*). For the same shunt loss as in the above example, this arrangement produces some 14% third harmonic and negligible second harmonic.

A further reduction in distortion can be effected by connecting a fixed resistance in series with the rectifier and driving the elements to a correspondingly lower a.c. resistance. This leads to the circuit of Fig. 6(*c*), which shows the arrangement applied directly as a transmitter shunt. The capacitor C is necessary in this case to prevent the direct voltage across the transmitter affecting the rectifier biasing, which is derived from the voltage drop in R_1 (the line current flowing through R_1 is little affected by changes in transmitter resistance because of the comparatively high resistance of the exchange transmission bridge). The capacitance *C* must be fairly high if amplitude/frequency distortion is to be avoided. However, the necessity for this capacitor can be avoided by circuit rearrangement.

The circuit of Fig. 6(*c*) is the basis of the regulator adopted for the type-700 telephone set.

(4.2.2.3) *Capacitance Effects.*

An undesirable characteristic of copper-oxide and selenium rectifier plates in this application is that of self-capacitance. This is an extremely variable quantity for any given type of rectifier plate, and varies with age, temperature and direct voltage applied. The shunting effect of this capacitance is most serious at low biasing voltages, when the forward resistance of the rectifier is high; for example, for a typical ¼ in diameter selenium rectifier disc the capacitance with a voltage of 0·3 volt applied in the forward direction is about 0·02 μF.

(4.3) Circuit for Optimum Performance

Disregarding for the present the number of additional components involved, an automatic sensitivity regulator having the desired characteristics might be of the form shown in Fig. 7,

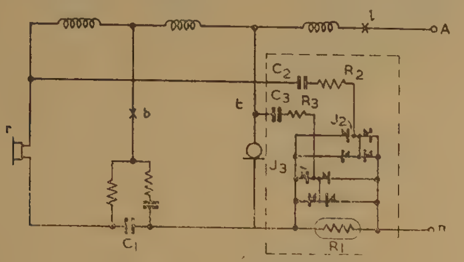


Fig. 7.—Regulator circuit for optimum performance.

where, on short lines, the rectifier assembly provides a.c. shunts directly across the transmitter, and via C_1 (the normal balance and d.c. blocking capacitor) across the receiver. Alternatively the same arrangement may be used to provide a line and balance shunt by connecting C_2 to *l* instead of *r*, and C_3 to *b* instead of *t*, the values of R_2 and R_3 being changed accordingly.

The non-linear resistor R_1 is chosen so that on long lines its resistance is low and the voltage drop across it is small, and the rectifiers are substantially unbiased. On short lines its resistance increases, and the direct voltage across it biases the rectifier paths so that they present low impedances; these impedances are much lower than the actual values required for shunting purposes, and R_2 and R_3 are chosen to limit the shunt loss accordingly. C_2 and C_3 prevent direct voltages within the telephone set affecting the rectifier biasing; their reactances must be low compared with the remainder of the shunting path to avoid frequency distortion. So that the circuit can be used with either direction of line polarity, parallel pairs of oppositely poled rectifier elements are necessary as shown.

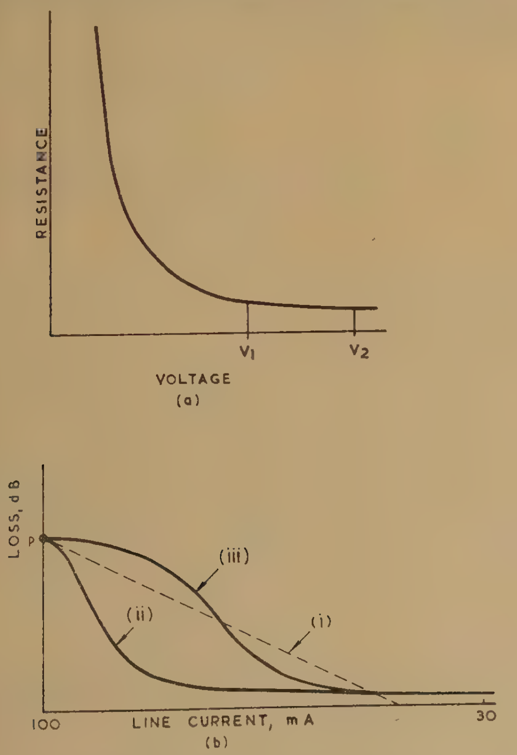


Fig. 8.—Effect of the choice of R_1 on the shape of the regulator sending-loss/line-current characteristic.

(*a*) Effect of bias voltage on the resistance of the rectifier element.
(*b*) Change of regulator loss with line current:
(i) Required sending loss.
(ii) Loss if R_1 is chosen to give bias V_1 at 100 mA.
(iii) Loss if R_1 is chosen to give bias V_2 at 100 mA.

Flexibility in shaping the regulator-loss/line-current characteristic is obtained from the choice of R_1. For example, suppose a shunt loss of *P* decibels is required at 100 mA line current, with negligible loss at 30 mA, as shown by (i) in Fig. 8(*b*). If R_1 is chosen so that the rectifiers are biased to the point V_1 on their resistance/voltage characteristic [Fig. 8(*a*)] with 100 mA line current, a small reduction in bias voltage will cause a rise in resistance and therefore a fall in regulator loss, and the loss/current curve will be similar to (ii) in Fig. 8. However, if a higher value of R_1 is chosen, so that, for 100 mA line current the rectifiers are biased at V_2, a small reduction in bias voltage will cause very little change in loss, and, in fact, the loss will not decrease appreciably until the voltage has fallen to V_1. Curve (iii) shows how such biasing maintains the regulator loss over a range of the higher line currents. This facility is especially important in a regulator for a universal telephone instrument requiring full loss not only on short direct exchange lines, but also for short extension lines on a parallel-feed p.b.x., where line currents are somewhat lower than on direct exchange lines.

Two minor disadvantages of the circuit are as follows:

(*a*) On long lines there is a small additional loss, of speech-clipping form, due to the unbiased rectifiers being in shunt across the transmitter; this can be minimized by suitable choice of rectifier elements and driving voltage.

(*b*) There is a small residual long-line loss due to the series element R_1.

A more serious practical drawback is the expense and bulk of the two capacitors C_2 and C_3.

(4.4) Practical Circuit

The need for the capacitors C_2 and C_3 can be avoided by using the more practical arrangement of Fig. 9, the regulator shunt

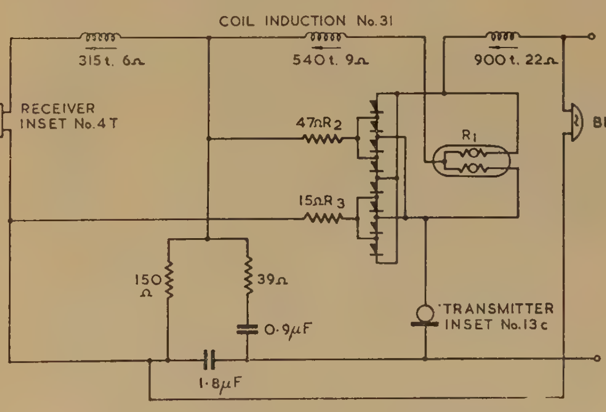


Fig. 9.—Transmission circuit of Telephone No. 706.

paths being connected across the transmitter via the d.c. blocking capacitor and across the 540-turn winding of the induction coil. The degradation in performance resulting from this modification is quite small.

The basic transmission circuit of the new telephone is shown in generalized form in Fig. 10, where Z_T, Z_R and Z_B represent the impedances of the transmitter, receiver and balance network, and Z_L is the impedance connected across the line terminals. Although this circuit at first appears to be very different from that of the 300-type telephone, it will be seen in Appendix 8 that actually the two circuits are very similar. The configuration of Fig. 10 is retained for the new telephone, because, as will be seen later, it lends itself more easily to regulation.

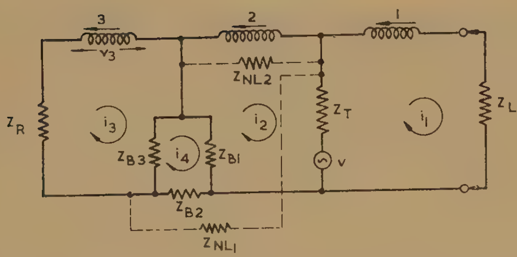


Fig. 10.—Generalized basic transmission circuit of Telephone No. 706.

Let the self-impedances of the windings be Z_1, Z_2 and Z_3, and the mutual impedances between the windings be Z_{12}, Z_{13} and Z_{23}, the subscript denoting the two windings concerned. Sidetone is then the power in the receiver due to an e.m.f., v, generated in the transmitter. Good sidetone suppression is obtained when the ratio i_3/v is low.

The voltage v_3 developed across winding 3 is the sum of three components, i_1Z_{13}, i_2Z_{23} and i_3Z_3. The equation for the (Z_R, Z_{B3}, winding 3) network is therefore

$$i_3(Z_{B3} + Z_R + Z_3) + i_1Z_{13} + i_2Z_{23} - i_4Z_{B3} = 0$$

Thus
$$i_3 = \frac{i_4Z_{B3} - (i_1Z_{13} + i_2Z_{23})}{Z_{B3} + Z_R + Z_3}$$

and clearly sidetone balance is obtained when the total induced voltage in winding 3 is equal and opposite to the voltage developed across Z_{B3}. For any condition other than that of balance, i_3 is determined by the above expression.

For reduction in sending sensitivity without the use of a capacitor (such as C_3 in Fig. 7), a shunt path may be connected across the transmitter as shown dotted via Z_{NL1}, which represents a variable impedance. The complete path is then Z_{NL1} in series with Z_{B2}, the latter being a low-reactance capacitor. However, this upsets a good sidetone balance condition, because, in practice, the voltage developed across Z_{B3} is reduced (owing to the reduction in i_4) more than v_3 falls by the reduction in transmitter potential difference. Balance can therefore be restored only if v_3 is reduced accordingly. This can conveniently be effected by shunting any of the windings, the obvious choice being winding 2, since this shunt circuit has a common connection with the transmitter shunt and needs no capacitor. This path is shown in Fig. 10 via a variable impedance Z_{NL2}.

Fig. 11 shows typically how the receiver voltage v_R is reduced

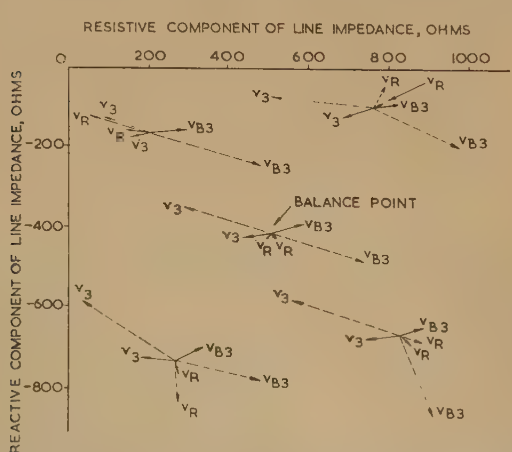


Fig. 11.—Reduction of sidetone by shunt paths Z_{NL1} and Z_{NL2} ($f = 1\,592$ c/s).
---- Without shunt paths.
—— With shunt paths.

by this method, even when the balance is poor, for any line impedance.

Now
$$v_R = -(v_3 + v_{B3})$$

where v_R and v_{B3} are the instantaneous voltages across Z_R and Z_{B3}, respectively. The vectors v_3 and v_{B3} are shown positioned at various line impedances, and for each value, the vector v_R has been obtained as above both for the non-regulated circuit (dotted vectors) and when Z_{NL1} and Z_{NL2} are added. In all cases the addition of these shunt paths reduces v_R except, of course, at balance, when theoretically v_R is already zero.

Z_{NL2} creates an additional power loss which affects both sending and receiving efficiencies; thus, by careful choice of Z_{NL1} and Z_{NL2}, so that not only are they in the correct ratio for a greatly improved sidetone balance but their respective sending and receiving losses total the value required, the circuit features of Fig. 7 can be obtained without the use of capacitors.

In the Telephone No. 706 the transmission circuit has been designed so that, not only is the ratio Z_{NL1}/Z_{NL2} equal to 0·52, the optimum for sidetone reduction, but also sending and receiving losses approximating to those required are obtained. This method is capable of introducing much greater sidetone attenuation when the regulator is fully operative than is obtainable with transmitter and receiver shunts, for which the maximum reduction is approximately equal to the total sending and receiving loss (in this case, 10 dB).

Applying the regulator circuit of Fig. 7 without the capacitors C_2 and C_3 directly to the circuit of Fig. 10, to provide the non-linear paths Z_{NL1} and Z_{NL2} as suggested, leads to the further complication that the two halves of each push-pull set of rectifiers are shunted by d.c. paths of different resistance, and the requirement of equal d.c. biasing of the rectifiers is no longer met. This is overcome by tapping winding 2 to the centre point of the current-sensing resistor, R_1, as shown in Fig. 9. When the circuit is redrawn as in Fig. 12 (in which one shunt path and the

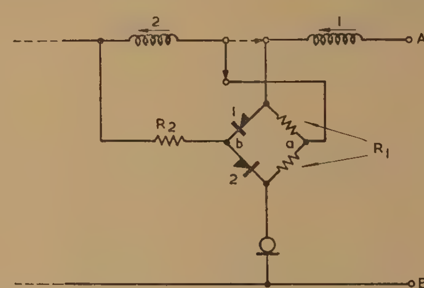


Fig. 12.—Centre-tapped current-sensing resistor.

oppositely-poled rectifier elements have been omitted for clarity) it is then obvious that, provided that the rectifier elements are similar, no voltage exists across a and b of the bridge, and d.c. paths between a and b have no effect on the rectifiers. Without the centre-tap, and leaving winding 2 connected directly to winding 1 as shown dotted, it is seen that R_2 and winding 2 form a d.c. shunt across rectifier 1 only, thus causing unbalance. The high d.c. resistance (about 1 kilohm) of the bell is not catered for in this arrangement and therefore does cause a small degree of unbalance, especially on long lines when this resistance is lower than that of the rectifier it shunts. The effect is negligible owing to the very low d.c. driving voltages involved.

The rectifier arrangement is such that all the plates may be conveniently assembled in series, with interconnections as shown in Fig. 9. A suitable characteristic is obtained from two ¼ in-diameter selenium plates in series for each half of a push-pull pair, as shown in Table 1.

Table 1

TWO ½ IN DIAMETER SELENIUM PLATES IN SERIES

V Direct voltage (forward)		Approximate current/voltage relationship $I = 1{\cdot}5V^n$
volts	mA	
0·7	0·13	$n = 7{\cdot}0$
1·0	1·6	$n = 6{\cdot}8$
1·3	8·0	$n = 6{\cdot}6$

The d.c. resistance of a pair of plates is about 12 kilohms with 0·6 volt applied, falling to 70 ohms with 1·5 volts applied. The whole assembly consists of a tube of 16 plates, tapped at every second one.

Consider now the circuit of Fig. 9 when connected by a subscriber's line to a 50-volt Stone transmission bridge with 200 + 200 ohm relays. Table 2 shows the approximate regulator operating conditions for the two extremes of line length:

Table 2

	Zero-loss line	1-kilohm line
Line current, mA	95*	31
Current through R_1, mA ..	76	31
Resistance of R_1, ohms	36	10
Rectifier driving voltage (each section), volts	1·35	0·15
Current through each forward rectifier path, mA	9·5	A few micro-amperes
Regulator network impedance between junction of R_2 or R_3 with the rectifiers and the centre point of R_1 at 1 kc/s	20 ohms (non-reactive)	3 770/70°†

* Initially 100 mA, falling gradually (owing to heating effects within the relay coils) to 90 mA after some 20 min.
† Equivalent to 11·5 kilohm resistance shunted by a capacitance of 0·04 μF.

On a 1-kilohm line the permissible sending loss of 0·3 dB suggested in Section 4.2.2 is just reached at the upper end of the frequency response, while at low frequencies the loss is negligible. A further source of long-line loss already mentioned is the slight clipping of the peaks of the speech waveform from the transmitter by the rectifiers, giving a total measured loss of about 0·3 dB on sending and less than 0·1 dB on receiving; this degree of clipping is not detectable by a listener.

The robustness of the resistance lamp is clearly an important consideration. Its filament is short, it is not of unduly fine gauge, and as it is normally run at little more than dull red heat, its life can be expected to be very long. Also, the parallel-connected rectifiers provide a measure of surge protection. In any case, failure of the lamp does not put the telephone completely out of action; fracture of either section of the filament puts the regulator permanently into the full-loss condition, and simultaneous fracture of both filaments also causes excessive sidetone but the telephone still remains usable.

Life tests have been carried out in the laboratory on 100 complete regulators running at line currents of 100–200 mA applied for 3 min periods separated by off-periods of 1 min; in addition, once per day the current was applied continuously for one hour. These were intended to simulate the most severe conditions likely to arise in normal use, and the regulators under test withstood the equivalent of 20 years' service in an average telephone without any failures.

(5) TELEPHONE No. 706

As already indicated, the final model of the new 700-type telephone incorporating the regulator is the Telephone No. 706. It is now in production and is shortly to be made available to telephone subscribers on the British telephone network. The circuit design, shown in Fig. 9, is essentially that of a complete regulated telephone rather than that of a conventional circuit to which regulation has been subsequently applied. By accepting the principle of sensitivity reduction on short lines, it becomes possible to design for maximum transmission efficiency on 1-kilohm lines, any efficiency impairments on short lines brought about by so doing being then adjusted for by the amount of regulation provided. For example, the core of the induction coil is gapped for highest inductance at about 30 rather than 100 mA, the result being slightly increased efficiency on long lines, but reduced efficiency and higher sidetone on short lines. The latter effect is covered by regulation.

The Y ratio of the circuit is 3·3, and the balance network is designed to use 10% preferred-value carbon resistors. The impedance naturally varies considerably with line current owing to the introduction of the a.c. shunt paths at high currents. On long lines the impedance is some 850 ohms with a slight positive angle. This falls to about 200 ohms (+15° to 30° depending on frequency) on short lines. These impedances are of a slightly better order than those of the current set in their effect on the sidetone performance of a distant telephone when working in the British Post Office network, both as it exists at present and also as it may be in the future when many of the 300-type sets have been replaced by the 700 type.

The instrument is housed in a new design of moulded case, more in accordance with present-day trends in styling (Fig. 13), and it will be available in a variety of colours.

Fig. 13.—Telephone No. 706.

(5.1) Transmission Performance

As shown in Fig. 3, the sensitivity control provided by the regulator is within ±1·0 dB of the optimum over the whole range of line lengths (with 600-ohm junction) and the overall transmission performance of the Telephone No. 706 is substantially better, both in loudness efficacy and in frequency response, than that of the Telephone No. 332 which it supersedes. Thus, on very short lines, on which the sensitivity of the Telephone No. 332 was slightly too high, the Telephone No. 706 is 2 dB quieter on sending, approximately equal on receiving and gives

some 4 dB less sidetone. On long lines it is 4 dB louder both on sending and on receiving, and it gives the same standard of transmission on a local line of 1 kilohm of 6½ lb cable as did the Telephone No. 332 on 660 ohms of 6½ lb cable.

A repetition of the field trial described in Section 3.1, carried out at one of the Telephone Manager's Offices with some 200 pre-production models of the Telephone No. 706, showed clearly that it was preferable on all counts, and that there was no serious body of complaints of excessive loudness on local calls.

The actual number of opinions expressed by the users at this office on this and the earlier trial were as shown in Table 3. It must be remembered in comparing these results that the trial of Telephone No. 706 followed immediately upon the trial of Telephone No. 700, and that the users were therefore, to some extent, conditioned to very loud reception.

Table 3

Opinions on Loudness of Received Speech on Internal Calls

Opinion	Telephone No. 700	Telephone No. 706
Deafening	12	1
Much too loud ..	32	—
Too loud	108	1
Satisfactory	15	135
Too quiet	—	14
Much too quiet ..	—	5
No opinion given ..	17	—
Total	184	156

Measurements of speech voltage on outgoing calls made immediately after substituting the Telephones No. 706 for the earlier unregulated type 700 showed a decrease in average speech voltage of 1·3 dB; here again (as observed in Section 3.1) changed sidetone conditions were undoubtedly having an effect. In this case, the quieter sidetone caused an increase in talkers' vocal level.

An interesting characteristic of the Telephone No. 706 is of practical value on short-line extension working, where two telephones in parallel are used simultaneously. Because of the sharing of line current between the two sets, the regulators allow the sensitivities of the instruments to increase, thus offsetting the fall in transmission performance normally experienced.

(5.2) Performance on P.B.X.'s

On direct exchange lines and on many types of p.b.x.'s the regulator meets all the requirements. In the case of some parallel-feed p.b.x.'s with exceptionally low line currents, less regulation is required on sending owing to the lower efficiency of the transmitter, but normal regulation of 4 dB is required on receiving for extension-to-extension calls with short lines. The regulator loss given in these cases is less than that required, the worst condition being that in which each extension telephone is fed with a current of about 40 mA, which is the highest line current at which the regulator loss is zero. If, as appears likely, telephones of the future will all include automatic regulation based on line-current sensing, it may be necessary in future p.b.x. design to ensure that the cord circuits provide a suitable line-current range.

(5.3) Flexibility of Design

The basic design of the regulator circuit is flexible, and to meet special requirements the regulator losses can be increased or decreased simply by lowering or raising R_2 and R_3 in Fig. accordingly. Some 16 dB total loss is obtainable when both R and R_3 are zero.

(6) ACKNOWLEDGMENTS

The development of the telephone was carried out jointly b engineers of the Post Office Engineer-in-Chief's Office and of th telephone industry, through the medium of the British Telephon Technical Development Committee, and is a happy example o fruitful co-operation between a Government Department an industry.

Acknowledgment is made to the Engineer-in-Chief of the Pos Office for permission to publish the paper. The authors' thank are particularly due to their colleagues, Messrs. D. L. Richard and R. B. Archbold, who have contributed much of the dat contained in the paper, and on whose work on the assessment o telephone transmission performance so much has been based and to Mr. H. J. C. Spencer, who, as liaison officer, has bee responsible for co-ordinating the development between the Pos Office and the manufacturers.

(7) REFERENCES

(1) Swaffield, J., and Richards, D. L.: 'Rating of Speec Links and Performance of Telephone Networks', *Proceedings I.E.E.*, Paper No. 2666, June, 1958 (**106 B**).

(2) Strecker, F., and Von Susani, G.: 'Investigations of th most Suitable Range of Reference Equivalent in Telephone Operating', *Elektrische Nachrichten Technik* **19**, p. 241.

(3) Inglis, A. H., and Tuffnell, W. L.: 'An Improved Telephone Set', *Bell System Technical Journal*, 1951, **30** p. 239.

(4) Bennett, A. F.: 'An Improved Circuit for the Telephon Set', *ibid.*, 1953, **32**, p. 611.

(5) Spencer, H. J. C., and Wilson, F. A.: 'The New 700-Typ Telephone', *Post Office Electrical Engineers' Journal* 1956, **49**, p. 69.

(6) Palmer, R. W., and Brimmer, W. L.: 'The British Telephone Technical Development Committee', *ibid.*, 194 **41**, p. 193.

(7) Roberton, J. S. P.: 'The Rocking-Armature Receiver *ibid.*, 1956, **49**, p. 40.

(8) Spencer, H. J. C.: 'Some Principles of Anti-Side-Ton Telephone Circuits', *ibid.*, 1956, **48**, p. 208.

(9) Richards, D. L., and Swaffield, J.: 'Assessment of Speec Communication Links', *Proceedings I.E.E.*, Paper No 2605, April, 1958 (**106 B**, p. 77).

(10) McMillan, D.: 'Some Performance Characteristics of th Subscriber's Telephone Transmitter', *Post Office Electrica Engineers' Journal*, 1935, **28**, p. 167.

(8) APPENDIX: SIMILARITY BETWEEN 700 AND 300-TYP TRANSMISSION CIRCUITS

Fig. 14(*a*) shows the basic transmission circuit used in th Telephone No. 706. In Fig. 14(*b*) the balance network Z_1, Z_2, Z is replaced by the equivalent T-network Z_A, Z_B, Z_C, where

$$Z_A = \frac{Z_1 Z_2}{Z_1 + Z_2 + Z_3}, \text{ etc.}$$

Fig. 14(*c*) is identical electrically with Fig. 14(*b*), and Fig. 14(*d* is obtained by interchanging the series components, winding

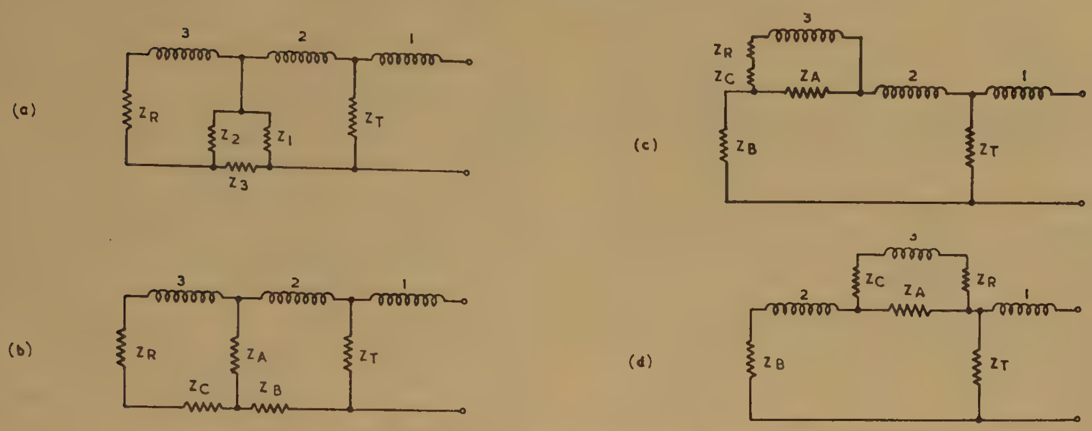


Fig. 14.—Equivalent transmission circuits.

and the (winding 3, Z_R, Z_C, Z_A) network and similarly winding 3 and Z_R.

Fig. 14(d) is identical with the 300-type circuit except for the additional component Z_C.

This is a demonstration of circuit configuration only. In practice, Z_A, Z_B or Z_C may not be physically realizable, in which case slight component changes are necessary to produce a practical circuit.

DISCUSSION ON 'TRANSISTOR ACTIVE FILTERS USING TWIN-T REJECTION NETWORKS'*

Mr. R. Hutchins (*communicated*): It is commonly believed that, if a twin-T network is used in the feedback path of a selective amplifier, the selective frequency is a function of the components in the twin-T only. I would like to point out that this is true only if there is no spurious phase displacement within the amplifier (other than 0 or 180°, depending on the design).

In the paper there is no mention of the phase error in the transistor amplifier, and in Appendix 9.2 it is implied that the frequency of maximum selectivity is at the balance frequency of the twin-T network. The phase shift of the amplifier caused by the current gain of individual transistors at 600 c/s as used in Fig. 3 is negligible. However, the second transistor, T_2, has a $10\,\mu F$ emitter by-pass capacitor whose reactance is about 27 ohms at this frequency. The effect of this is to add a substantial reactive term to the input impedance of the second transistor, the value of the input impedance being approximately

$$Z_{in} = r_b + \frac{r_e}{(1-\alpha)} - j\frac{27}{(1-\alpha)}$$

Hence, the current supplied to this transistor leads the output current of the previous transistor by an angle of about 10°, depending on the parameters of the transistor used.

The effect of this on the performance of a selective amplifier is to cause the feedback to be just positive on one side of the null frequency, so that the optimum gain of the amplifier is just off the null frequency. This argument may also be substantiated by evaluating the input impedance to the complete amplifier, whose variation with frequency is the true cause of the selectivity. The minimum input impedance is then found to be on one side of the null frequency of the twin-T, although at the condition for minimum the input impedance is not resistive.

It follows that the phase shift in the amplifier causes an optimum response which is not at the null of the twin-T network. However, if the gain of the amplifier is high, as assumed in the paper, the error caused by the phase shift will be very small because the quadrature component of the transmission locus of the T-network changes very rapidly with frequency.

Mr. A. E. Bachmann (*in reply*): I am glad that Mr. Hutchins points out the influence of phase shift in the amplifying stages on the performance of the entire selective amplifier.

* BACHMANN, A. E.: Paper No. 2787 E, March, 1959 (see 106 B, p. 170).

621.385.1.032.213.6

The Institution of Electrical Enginee
Paper No. 2750
Nov. 19

HIGH-CURRENT-DENSITY THERMIONIC EMITTERS: A SURVEY

By A. H. W. BECK, M.A., B.Sc.(Eng.), Associate Member.

(*The paper was first received* 8*th January, and in revised form* 11*th August*, 1958. *It was published in November*, 1958, *and was read before th* RADIO AND TELECOMMUNICATION SECTION 19*th January*, 1959.)

SUMMARY

The steps leading to the development of modern emitters are described and the characteristics and operation of such emitters are discussed.

(1) INTRODUCTION

Thermionic emitters were last surveyed by Wright[1] before The Institution in 1953. His work still stands as a valid statement of our knowledge of the emitting systems which he treated.* Progress since that date has mainly been in two directions, the improvement of our understanding of the behaviour of oxide cathodes under pulse conditions, and the introduction of improved types of emitter, especially those intended to furnish high values of continuous current density. It is with this field that the present survey is concerned.

I have written this paper with three main objectives in mind. First, to attempt to give a connected account of the manner in which the types of cathode in use to-day were evolved, and of their relation to earlier knowledge. Secondly, to give an account of the characteristics of such cathodes, emphasizing the specific advantages and disadvantages so that prospective users will be able to choose the cathode best suited to their needs. Thirdly, to provide an account of the present state of our knowledge, in the hope of stimulating further fundamental research on these interesting and useful devices.

(2) WHY HIGH-CURRENT-DENSITY CATHODES ARE REQUIRED

As a preliminary to the discussion of the properties and characteristics of modern high-density cathodes it is necessary to know why such cathodes are needed. In turn, this requires a little knowledge of the limitations of oxide cathodes.

Oxide cathodes are generally operated at temperatures between 900° K and 1 100° K. If they are operated at temperatures below the minimum quoted, their life is short, which is said to be due to poisoning of the cathode by residual gas in the valve. At temperatures above the maximum quoted, life is again short, because of gross evaporation of the cathode coating. Reference to tabulated properties of oxide cathodes shows that emissions of several amperes per square centimetre should be available, but if one makes the experiment, it is found impossible to dra much more than 1 A/cm^2 from the cathode, because the coatin overheats and vaporizes. This is due to Joule heating in th series resistance of the cathode coating. At 1 A/cm^2 this become roughly equal to the power density taken from the heater t raise the emitter to the operating temperature, and if the emissio density is raised above this value the system becomes unstabl Of course, there are ways in which this can be improved, th most obvious being to work with very thin layers of coating but this is objectionable from the viewpoint of life. Thus ther is a limit somewhat less than 1 A/cm^2 above which one cann operate conventional oxide cathodes. However, up to th limit there is no evidence to suggest that increased curren density *per se* has a bad effect on life, and a cathode operatin at 0·5 A/cm^2 in a good environment can live much longer tha one operating at 0·05 A/cm^2 in a poor one.

Normal receiver valves in fact work at low current densities c around 0·05 A/cm^2. The factor of safety between this valu and the saturation emission of the cathode covers variation between the initial activities of the cathodes in different valve and the decrease of the cathode activity during life.

The provision of a cathode capable of a high emission densit in such a valve would only provide for a larger factor of safet under the second head, and therefore a longer life.

In low-frequency transmitter valves the total cathode curren required may be very high, but here there is no particula objection to increasing the cathode area so that the densit remains tolerable, since the dimensions of the valve can b increased without impairing its efficiency.

At high frequencies and especially at microwave frequencie matters are very different because the dimensions of the valv are now determined in some way by the operating frequenc To discuss this limitation we must introduce the concept c scaling. Suppose that we have been able to design an acceptab valve to operate at some frequency f_1. How do we change th dimensions so as to obtain comparable performance at a highe frequency f_2? Two types of scaling are possible. In the firs known as complete scaling, we show by dimensional analysis c Maxwell's equations that these equations are invariant under transformation which leaves the voltage constant and transform the dimensions by the frequency ratio, i.e. a dimension l_1 at transforms to ml_1 at f_2 where $m = f_2/f_1$. It is clear that in th type of scaling the current density varies as m^2 and doublin the frequency quadruples the required cathode density. T illustrate what this can lead to, we note that a density of 1 A/cm at 200 Gc/s scales to a density of 1 μA/cm^2 at 200 Mc/s, a valu which we can take as representing the present limiting frequenc for conventional receiving valves using grids. One could n design a very good valve with such a current density.

Some improvement is obtained if we use the second and le stringent form of scaling, in which only the transit angles remai invariant. Here the dimensions scale as $V^{1/2}m^{-1/2}$ and th current density scales as $V^{1/2}m^2$, so that if the voltage is reduce the density ratio is not so unfavourable. Unfortunately, th power input scales as $V^{5/2}$ so that $V^{1/2}$ cannot be too muc reduced.

* It is unfortunate that the article on 'Thermionic Emission'[2] prepared by W. B. Nottingham in the new 'Handbuch der Physik' is highly idiosyncratic in two respects. First, it almost ignores the mobile donor hypothesis of Nergaard and his co-workers, and secondly, it puts much weight on Nottingham's hypothesis that many of the phenomena observed in real diodes are only explicable in terms of a high reflection coefficient from the emitting boundary for slow electrons, a consideration which also applies to the anode of the retarding-field diode. This reflection coefficient has to be much greater than the 5–10% effect predicted by theory, and the direct support for it is an experiment by Hutson[3] on a single crystal of tungsten. However, more recent and more direct measurements on single-crystal tantalum by several investigators show no such effect; see for example Shelton.[4] Wright's remarks on the error of supposing that reflection effects are important seem substantiated.

This is an 'integrating' paper. Members are invited to submit papers in this category, giving the full perspective of the developments leading to the present practice in a particular part of one of the branches of electrical science.

Mr. Beck was with Standard Telecommunication Laboratories, Ltd., and is now at the Engineering Laboratory, Cambridge University.

In the face of these considerations it might be wondered how millimetre valves have ever been made to work. We have so far neglected one factor. We have tacitly assumed that the current density in the valve is actually that emerging from the cathode. Luckily, in microwave valves using focused beams this is not the case, and the cathode density may be much lower than that in the beam. In the well-known Pierce gun the optimum ratio of densities is calculable and turns out[5] to be 26. For some empirically designed guns[6] the ratio is higher; certainly as high as 75, and figures of around 200 have been claimed. Let us assume that we gain a factor of 100 through this means. Then our earlier calculation takes us to a figure of $0 \cdot 1\,\mathrm{mA/cm^2}$ at 200 Mc/s. Using the best high-density cathodes we can increase the density to $10\,\mathrm{A/cm^2}$ at 200 Gc/s so that our final best result to-day corresponds with a figure of $1\,\mathrm{mA/cm^2}$ for our low-frequency valve. It has to be admitted that this is not an impressive figure: pulse operation gives us another factor of 10. These considerations explain rather dramatically why one is interested in improving cathodes.

If one takes the emission constant usually quoted for oxide cathodes and works out the emission available at around 1 100° K, i.e. at the maximum operating temperature, one finds that values of several amperes per square centimetre are obtained. We discuss the main reason why this is not useful in the next Section.

(3) THE HEAT BALANCE IN A CATHODE

Here we consider that sufficient current is drawn from the cathode to run it just into saturation. This is a reasonable assumption if we are interested in evaluating the maximum usable emission. The heat balance is then an equilibrium between two sources of heat and two sinks of heat, namely the heat supplied by the heater, the Joule loss in the coating, the radiation from the cathode surface and the power given to emitted electrons. When no emission is drawn from the cathode, an equilibrium is established between the heater power and the radiation.

Then

$$P_H = C_T T^n \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where C_T in general is a function of temperature and n is a number approximately 4. When the saturation current I_s is drawn, we have

$$P_H + I_s^2 R_s = C_T T^n + I_s\left(\phi + \frac{2kT}{e}\right) \quad . \quad . \quad . \quad (2)$$

for equilibrium. Here the new terms are the Joule loss and the power lost to the electrons, which is in two parts, that required for the electrons to surmount the work-function barrier and that corresponding to the initial electron energies, assumed to follow a Maxwellian distribution. Now consider that the heater input is not changed as current is drawn. Then, for equilibrium,

$$I_s = \frac{1}{R_s}\left(\phi + \frac{2kT}{e}\right) \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and the cathode will violently overheat if I_s is increased above the value given by this equation. Moreover, because of the square law in eqn. (2), the increase in heating power is very rapid once the limit is passed. If, for an oxide cathode, $I_s = 1\,\mathrm{A/cm^2}$ at 1 100° K, eqn. (3) shows that $R_s \simeq 1 \cdot 6$ ohm/cm². For this temperature $P_H \simeq 3$ watts/cm². For $I_s = 2\,\mathrm{A/cm^2}$ the Joule input is $6 \cdot 4$ watts/cm² and the temperature will increase by over 20%.

In dispenser cathodes the equilibrium represented by eqn. (3) occurs at much higher values of I_s, for example $6\,\mathrm{A/cm^2}$. Thus R_s must be $0 \cdot 25$–$0 \cdot 30$ ohm/cm² in this case. Very high c.w. emissions, of the order of $800\,\mathrm{A/cm^2}$, have been claimed for certain cathodes particulars of whose manufacture have not been disclosed.[7] It seems likely that, if the claims are substantiated, temperature-enhanced field emission is being observed rather than thermionic emission, for the considerations just discussed would lead to a very unlikely state of affairs.

(4) PULSE DECAY IN OXIDE CATHODES

The other great practical defect of oxide cathodes is the decay observed when emission is drawn during a long pulse. As is well known, very high emissions in excess of $100\,\mathrm{A/cm^2}$ can be drawn during pulses whose duration does not exceed about 10 microsec. If pulses whose duration is a few milliseconds are applied, a decay curve is observed in which the emission drops by a large factor, of the order of 10, depending on the precise temperature and other conditions. The exact mechanism of this effect is still doubtful. Some workers hold that it is entirely due to poisoning of the cathode by gas released from the anode under electron bombardment. Recently, however, the school of Nergaard[8] have given strong evidence for what is called the mobile-donor hypothesis. The hypothesis is that the active centres, from which the emission originates, are mobile under the influence of fields set up by the flow of current in the semiconductor and move away from the surface, leaving a high-resistance surface layer. Consequently the emission decays. Naturally both effects may, and probably do, coexist. Decay phenomena have been observed in modern dispenser cathodes, but their magnitude is less than in oxide cathodes. We should not too readily assume that the phenomenon is due to the same cause in both cases.

(5) LIFE IN OXIDE CATHODES

The life of oxide cathodes has been examined in great detail in recent years, especially by the workers at the Bell Telephone Laboratories and the Post Office Research Laboratories.[9, 10] For our present purpose it is unnecessary to describe more than the conclusions of this work, which stripped to essentials, are the following:

(*a*) The major causes of cathode failure encountered in ordinary commercial receiver valves are poisoning due to gas and the build-up of a resistive interface layer between the base-metal sleeve and the emitting oxide layer. The capacitance shunting this layer is small and therefore this resistance introduces degeneration, which is observed as a reduction in the effective mutual conductance. The growth of the resistive layer, which is aided by certain conditions of operation, e.g. high temperature or operation with the anode current cut off, depends in a complicated way on the impurities in the cathode sleeve, these being sometimes introduced to assist in the cathode activation.

(*b*) If interface resistance can be eliminated, as is possible by the use of very pure nickel or platinum sleeves,[10] then life becomes mainly a function of the gas situation in the tube. Gas may be present as a result either of incomplete pumping or of desorption under electron bombardment. If, by careful pumping and extra processing of the piece-parts, gas can be kept at a sufficiently low level throughout life, very extended lives in excess of 50 000 h can be obtained. The only remaining difficulty is thus to obtain a reasonable value of initial emission, which, with the very pure sleeves involved, may call for revised activation techniques.

Careful experiments show that there is no specific influence of current drain on life, provided, of course, that the gas pressure does not rise as a consequence of the greater current. On the contrary, in the absence of reducing impurities, electrolytic processes due to the passage of current assist in maintaining the cathode activity.

The dispenser cathodes obtain long life by a quite different mechanism. Throughout their life thè emitting surface is maintained by replacement from the interior of the cathode, where a vast reservoir of active material can be established. Provided the rate of arrival of active material at the surface is large enough, operation even in quite high gas pressures does not destroy the cathode. We are not able to say precisely what factors control the life, but we do know that heater life is a much more serious problem than is cathode life.

(6) POSSIBLE TYPES OF HIGH-EMISSION CATHODE

We now discuss the general characteristics of known high-density cathodes, and here it is convenient to define 'high density' as meaning 'capable of giving a continuous emission density above 1 A/cm^2 for a life in excess of several hundred hours'. Fig. 1 shows saturated emission plotted against temperature for

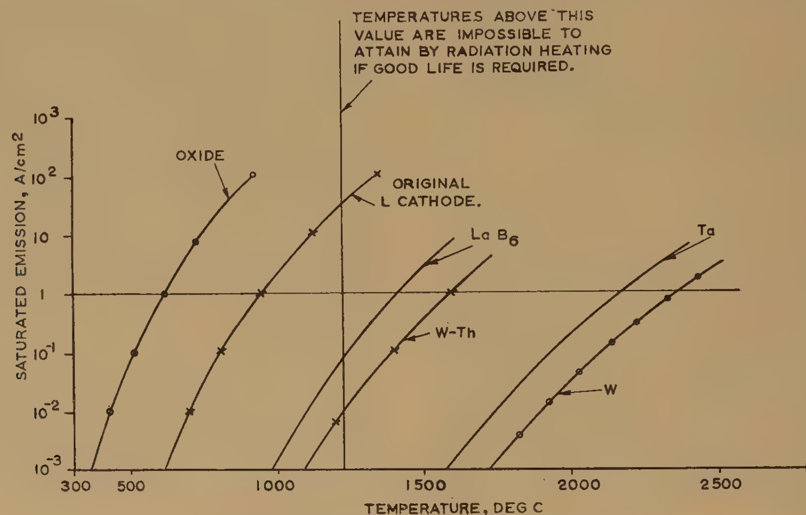


Fig. 1.—Emission of various materials.

a series of types of cathode. At least two pure metals, tungsten and tantalum and possibly two more, niobium and hafnium, must be considered. These all require to operate at very high temperatures. Next, we must consider thoriated tungsten and bulk thoria, both of which operate at temperatures in the range 1 500–1 600° C. Then we select the best of the alkaline-earth borides, investigated by Lafferty.[11] This is lanthanum boride and it gives high densities for temperatures above 1 400° C. The original L-cathode[12] is shown next, and here there is a sharp reduction to temperatures in excess of 950° C. Finally, the oxide-coated cathode represents the limiting low-temperature case, just coming up to our definition at about 600° C.

We first ask what considerations limit the use of pure-metal emitters, and we arrive at conclusions which are perhaps a little unexpected. Let us consider two valves both giving 10 kW unmodulated r.f. output. The first is a typical transmitting triode using a thoriated-tungsten filament, and the other a klystron for frequencies of a few thousand megacycles per second. The triode 3J/192E filament gives a peak current of 12 amp for a heating power of 330 watts. The working point is 7 kV, 2·6 amp. A rather similar valve with a tungsten filament (CV2687) requires 1·375 kW of filament power. The difference in filament power has a marginal influence on the overall efficiency and running cost. The real reasons why the thoriated filament is preferable are subsidiary ones, among which are the easing of the lead-in problem, reduced magnetic fields and therefore reduced hum and buckling forces, and less capital expenditure in the filament supply. The first of these reasons is rather impor[illegible] fo since the life of a tungsten-wire emitter is directly proportiona to the diameter, life can be increased at the expense of highe filament current and lower filament voltage, the limiting con dition being when the filament is reduced to a single, long, thin walled hollow cylinder. Such a cathode would require a ver low-voltage high-current supply and would present a difficu and expensive but not impossible lead-in problem.

In the klystron, the working point might well be 20 kV, 2 am Suppose we decide to obtain this from a tungsten disc. Allow ing a factor of safety of 2, we should need 4 cm^2 area at 2 360° C which would require around 700 watts of heating power, which in comparison with 40 kW, is negligible. The life would depen on the disc thickness, which scarcely affects the heating powe Modern vapour pressure data (Honig)[13] give 10^{-6} mm Hg fo tungsten at the quoted temperature, which is not excessivel high. The only difficulty is a practical one; the disc has to b heated by electron bombardment plus some radiation from th bombarding source. This involves the use of a separate highly insulated power supply and a significant overall increase in th power required. However, this system has been used in practic and has worked well, the only objection being the rather hig first cost. Consider now what would happen if we used thoriated-tungsten disc. We should probably like a somewha greater reserve of emission, so let us keep 4 cm^2 area and choos a temperature giving 2 A/cm^2, i.e. about 1 670° C; the inpu required drops to about 240 watts. However, very little gained because the disc cannot be raised to the required tem perature except by electron bombardment. This follows fro the work of Danforth and Haddad,[14] which shows that th heater would have to run at about 2 300° C to get the disc u to the required temperature by radiation. Thus, if this solutio were adopted, we should transfer the life problem from th cathode to the heater. We conclude, then, that the use o thoriated tungsten would not cheapen the installation sufficientl to make its use attractive.

We can now come to a rather useful conclusion, which is tha it is no use employing a lower-temperature emitter unless th temperature is lowered sufficiently to allow one to use the stan dard indirectly-heated cathode structure, operating with th heater temperature low enough for long life. An upper lim to the operation of ordinary alumina-covered heater wire often held to be around 1 800° K, which limits one to cathod temperatures below 1 500° K, or 1 227° C. At this point I a

writing with life figures in excess of about 3 000 h in mind. There are special applications where much shorter lives are tolerable, but these must be specially studied. This qualification disposes of the possible use of bulk thoria, which has been shown[15] to be a reasonable pulse emitter but a short-life continuous emitter. We are thus led to the consideration of only two groups of cathodes. First, we have to consider methods of improving the life of oxide cathodes at high temperatures and of modifying the oxide cathode so as to improve this feature of its performance. Secondly, we consider the modern dispenser cathodes, such as the L-cathode and its variants and the bariated-nickel or 'b.n.' cathode together with its variants.

(7) MEASUREMENT OF CATHODE EMISSION

Before starting the study of high-emission-density cathodes, it is necessary to discuss the difficulty of actually measuring the saturated emission from a good cathode in the operating range. The measurement technique can give rise to very serious differences in emission and can lead to differences between the results of different laboratories. We here eliminate low-temperature low-field techniques as they bear such a remote relation to the conditions under which our cathodes are used. The raw data are a series of values of observed cathode current as a function of anode voltage taken for several values of temperature. The temperature measurement is carried out either by optical pyrometry or by the use of thermocouples and is inherently more accurate than in the case of oxide cathodes. The reason in the first case is that the spectral emissivity is higher and very much more constant than for an oxide layer. The correction is then smaller and not subject to change. In the second case we can easily attach the thermocouple and we are dealing with a relatively massive metal body so that considerations of temperature drop across a thin layer need not influence us.

Some of the common ways of treating the raw data are as follows.

(7.1) The Schottky Plot

Here we plot $\log I$ against $V_a^{1/2}$ (more accurately against $E_a^{1/2}$, where E_a is the field at the cathode surface) going to values of V_a sufficiently great to ensure that a straight line is obtained. The straight line is then extrapolated back to $V_a = 0$, to obtain a value of zero field emission. The main objection to this procedure is that the slope of the Schottky line should yield the cathode temperature but never does. Processes not considered in the theory are therefore operating and one cannot say without further analysis whether the extrapolation is meaningful in a given case. Patch theory[16] shows clearly that there are certainly cases where the extrapolation is not meaningful.

A modification consists in plotting $\log I$ against V_a and extrapolating once more. There is no theoretical justification for this procedure, but it is just as good a relative measure as that described earlier.

These methods share the same major practical defect, namely the difficulty of taking a good cathode really far into saturation and the likelihood of spoiling it. This consideration basically limits the method to short pulses.

(7.2) Space-Charge Break Method

Here the data are plotted either on two-thirds-power-law paper or on log-log paper. In either case the plot is a straight line in the space-charge-limited regime, and the value of current at which the emission first falls appreciably below this line is taken as the zero field saturation current.

In a variation of this method the point of inflection in the I/V plot is determined by electrical differentiation. Until relatively recently this has been looked down on by the purists as lacking theoretical justification. However, Crowell[17] has now demonstrated that the inflection does in fact give a good measure of the zero field emission from an ideal emitter so that the theoretical foundation is now at least as firm as that for the Schottky plot.

The big advantage of these methods is that the valve does not have to be run as far into saturation and therefore d.c. measurements become more feasible, but there are certain practical difficulties. If the graphical method is employed it is difficult to define the break, whereas the inflection method demands the use of fairly complex circuit design.

Another modification of the method has, I believe, been in use at the Bell Telephone Laboratories for many years. It is illustrated in Fig. 2. Here the solid line represents the measured

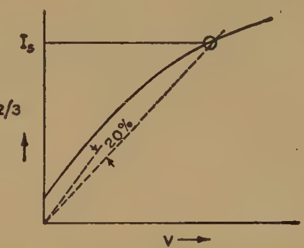


Fig. 2.—Bell Telephone Laboratories emission test.

$I^{2/3}/V_a$ plot for a real valve with finite emission velocities. The saturation emission is measured by drawing a line parallel to the straight initial part, but going through the origin, and then determining a line through the origin with 20% less slope. The intersection of the last line and the measured characteristic gives I_s. Here the cut is quite definite.

I have been able to give a theoretical justification for a rather similar method. In this the measured characteristic is plotted as before and the intersection with the straight line representing Child's law for the particular diode is used to define I_s. Fig. 3 illustrates the method. Here one must know the diode spacing to calculate the Child's law line. However, it can be shown that, provided the anode voltage for intersection is above 10 volts, the intersection gives I_s to a high degree of accuracy. If desired, a correction involving V_s (the voltage corresponding to I_s) and T can be included to improve matters still further. The proof of these statements depends on the solution of the diode equations in parametric form and not on the use of Langmuir's solution.

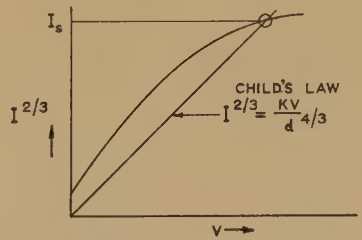


Fig. 3.—A modified method for emission testing.

Table 1 gives results for the saturation currents for both oxide and b.n. cathodes obtained by different measuring techniques. It fortunately turns out that the differences between methods are less for b.n. than for oxide cathodes, partly owing to the freedom from sparking and anomalous Schottky effect which is observed with these cathodes.

A point which is often overlooked in cathode evaluation

Table 1

SATURATION CURRENTS OBTAINED BY DIFFERENT METHODS

Type of cathode	T	Pulse emission, A/cm^2			
		Schottky plot	$\log I/V_a$	$I^{2/3}/V_a$	$\log I/\log V_a$
	deg C				
Oxide	800	7·0	4·2	5·7	7·5
	850	9·0	5·0	8·3	10·5
Bariated nickel	800	4·5	6·0	4·5	5·0
	850	6·0	8·0	6·0	8·0

studies is the importance of carrying out the measurement on a truly planar or truly cylindrical system.* In the latter case guard-ring diodes should be used. In the former, one too often sees diodes in which the emitting surface is the closed end of a cylinder of relatively small diameter, which contains the heater. This assembly is placed facing a large disc anode. Field concentration at the edges of the cathode gives rise to serious errors in the direction of increased emission.

All measurements subsequently reported on cathodes manufactured by the company are made in a properly designed diode in which the equipotentials are planar (Fig. 4). In the case of

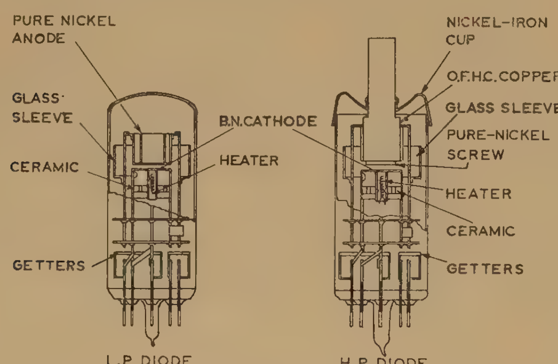


Fig. 4.—Standard low-power and high-power diodes.

measurements quoted from other authors' work it is often not known whether this requirement is fulfilled.

Another fruitful source of confusion involves the quotation of temperatures. The convention adopted here is to quote uncorrected optical pyrometer readings as degrees brightness (°B); corrected for spectral emissivity these become degrees centigrade, which usually involves the addition of 70–100° to the reading for the values most frequently used here. Unfortunately, some authors do not make the necessary distinction between brightness and centigrade temperatures.

A final point, which is appropriately mentioned here, is that some researches have been published on types of dispenser cathode which are known to have good performance, in which the investigators have found abnormally low values of emission. Such work has been disregarded in the sequel, for one can only conclude that the technique of preparation was faulty in some way. It is very unlikely that measurements on faulty cathodes have relevance to the behaviour of good cathodes, especially when the faults are due to the inclusion of indeterminate poisoning impurities.

* Nottingham (*op. cit.*) does not discuss this aspect of the Bell 1949 diode, which he says 'so nearly approaches the ideal structure'. From the illustration this diode would appear to be liable to edge effects, but possibly the Bell investigators have studied the question and are satisfied with the result.

(8) IMPROVING THE OXIDE CATHODE

Historically, the first requirement for improved performance from oxide cathodes was recognized in pulsed valves for wartime radar and especially in magnetrons. In both cases the difficulty was not so much that the emission was insufficient as the fact the valves sparked over when the applied voltage reached a certain level. It is, of course, well known that observations of these phenomena first disclosed the very high short-pulse emission of conventional oxide cathodes. Thus the first work directed towards high-performance oxide cathodes was aimed to increase the electric and thermal conductivities of the coating, to bond the coating more tightly to the core and to provide a greater volume of coating. This last aim had been earlier fulfilled by the original dispenser cathode of Hull,[18] but this was well adapted for gas-discharge devices and only poorly so for valves and magnetrons.*

Cathodes were therefore made in which nickel mesh was welded to the base sleeve, the interstices of the mesh being tightly packed with carbonate. Another variation was to plate a spongy mass of nickel on the sleeve and to fill this with active material. The reader should consult Collins[19] and Fisk, Hagstrum and Hartman[20] for details of these expedients. It soon became clear, however, that the problems of the magnetron were very specialized, as the thermionic emission is only required to start the oscillation, after which secondary emission takes over. Moreover, emission phenomena are obscured by the effects of back bombardment by electrons emitted with wrong phases. In spite of this back bombardment, the mesh and mush cathodes constituted a partial answer to the sparking problem and a step on the road to modern dispenser cathodes.

By the end of the war, work had started, especially at the Bartol Research foundation under Danforth,[15] on the properties of ceramic cathodes based on thoria. To say that this work at present seems to have lost some of its commercial importance is not to say it has lost any of its scientific interest and value. At the same time, work on modified oxide cathodes has continued. As an example we cite the work described by Huber and Charles.[21] These authors studied several modified oxide cathodes among which were:

(*a*) *A cathode formed by stacking short nickel tubes, whose diameter was a few tenths of a millimetre and whose length was about* 1 *mm, side by side.*—The interstices were packed with paste. The tubes were fixed by sintering with nickel powder. Clearly

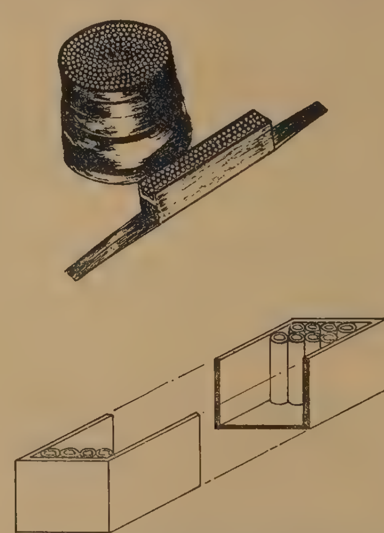


Fig. 5.—Huber's cathode with tubes.

* In view of later work on the L-cathode it should be noted that Hull used, *inter alia*, a form of barium-aluminate filling.

this cathode is merely an extension of the mesh cathode in which a greater reservoir of active material is provided. Results are quoted which show freedom from sparking up to 40 A/cm^2. Continuous densities of 500–600 mA/cm^2 could be drawn so that the performance is similar to that of a conventional oxide cathode. Fig. 5 shows this type of cathode.

(*b*) *Laminated cathodes.*—The coating in this case is applied to laminations stacked with their width parallel to the direction of electron flow. The laminations are spaced from one another by an amount equal to about half their width. The cathode is sprayed, taking care that the wide face is coated. The idea here is to shield the coating from ion bombardment. The best continuous currents were similar to those obtained from the above mentioned cathodes, and the cathode is obviously difficult

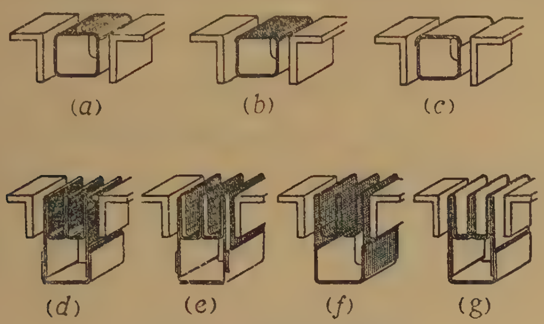


Fig. 6.—Laminated cathode.

to incorporate in electron beam devices. Fig. 6 shows several versions.

(*c*) *Faggot cathode.*—This is really a modified L-cathode in which the porous tungsten disc is made by swageing a mass of tungsten wires forced into a molybdenum tube. It therefore need not detain us.

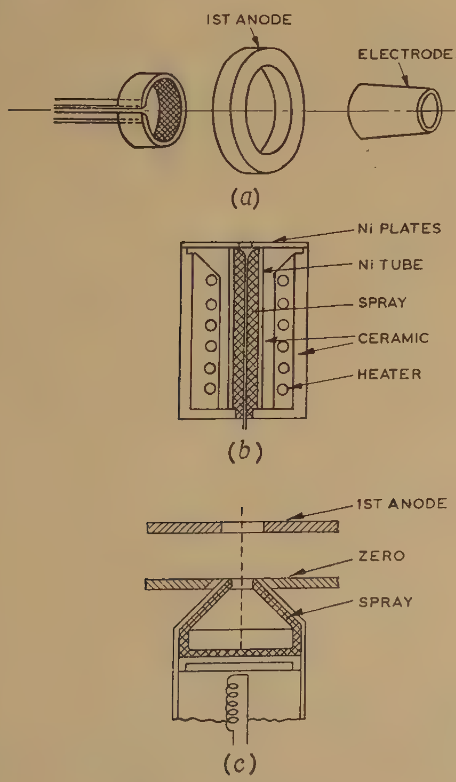


Fig. 7.—Hole cathodes.

(*d*) '*Hole*' *cathodes.*—Another advance in this field, which has excited considerable interest recently, is the 'hole' cathode. This actually has a long history; the version shown in Fig. 7(*a*) is described in Bruche and Scherzer,[22] where it is attributed to Johnson.[23] Espe and Knoll[24] show a version [Fig. 7(*b*)] patented by Schroter[25] which exhibits all the essential features of the modern ones. Recent investigations seem to have been provoked by an investigation by Babcock, Holshouser and Von Foerster,[26] who studied a spherical system and found that the I/V characteristic was quite different from that of a planar diode with a cathode area equal to the hole area, being much steeper and showing very little sign of saturation.

Further studies by Mueller[27] showed that the emission is not in fact drawn uniformly from the area of the hole but is concentrated on the edges of the hole giving rise to a hollow beam. The model now advanced for the operation is that barium-strontium alloy migrates from the interior of the box along the relatively thick edges of the hole. The emission is high partly because of the relatively high temperature of the metal face and partly because of the strong electric field. Owing, presumably, to the high operating temperature the lives of hole cathodes are short. Anderson and McEwan[28] found an empirical relation for the current as a function of voltage, temperature, hole diameter and spacing. They also found it was necessary to shape the cathode as shown in Fig. 7(*c*) if the beam was to be focused by immersing the cathode in a magnetic field.

(*e*) *Sandwich cathodes.*—A device very similar to the hole cathode has recently been described by Sugata and Nakamura.[29] This is shown in Fig. 8 and consists of a conventional oxide

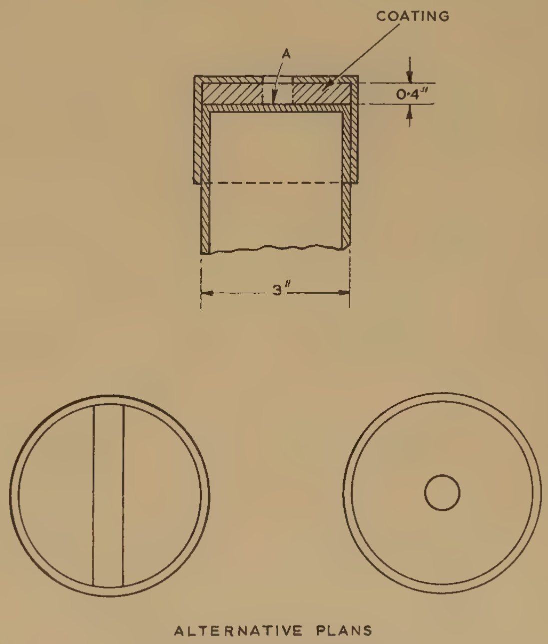


Fig. 8.—Sandwich cathode.

coating sandwiched between two cups, the upper of which is pierced by a hole. The distance from the top of the outer cup to the top of the inner cup is roughly equal to the hole diameter, so that emission can possibly be drawn from the face A as well as from the lips.

The emission distribution was measured by traversing a slit across the cathode face. The slit dimension was about one-third of the width of the cathode aperture, so that the resolving power was rather low.

The observations show

(i) Emission is detectable out to three hole diameters from the edge of the hole, but the magnitude is four below the value over the hole.

(ii) There is some slight evidence for a small fall in density when the slit is over the hole centre.

Electron optical studies showed that the effect of overheating was to produce a hollow beam, because the active material was evaporated from face A.

The sandwich cathode thus gives an extra design parameter which can be adjusted, at least partially, to homogenize the beam.

Interest in hole cathodes and the like does not seem to have been maintained. This is presumably because of short life and the difficulty of utilizing the cathode in beam tubes.

(9) THE EVOLUTION OF THE MODERN DISPENSER CATHODE

Rather surprisingly, the modern dispenser cathode has a quite lengthy history. Espe and Knoll[24] have a section entitled 'Barium sintered cathodes' in which they refer to patents taken out by Nienhold[30] (1924), Pirani and Ewest[31] (1928) and Espe and Evers[32] (1934). The description given by Espe and Knoll is very significant, as they speak of preparing a highly porous tungsten piece, soaking it in a solution of barium salts and converting to oxide by pre-heating in vacuum. The utilization of these cathodes seems to have been restricted to mercury-vapour lamps, and nothing is said as to their emission capabilities.

The next major publication in the field is the well-known dispenser cathode of Hull,[18] again proposed for use in gas-discharge tubes. Here we have the basic idea of continuous replacement of the active layer, which is pictured as a thin-film atomic emitter. The active material was in the form of grains of fused 'barya-alumina eutectic 70% BaO, 30% Al_2O_3 by weight'. This material was used to fill a closely-woven mesh cylinder of molybdenum wire which was heated by passing current through the wires. In the range 1 150–1 200° C sufficient reaction occurred to maintain the emitting surfaces, which were thin radial molybdenum vanes, with a covering of 'BaO and of reaction products with Mo' throughout life. Hull quoted figures for the emission constants of $A = 0{\cdot}85$, $\phi = 1{\cdot}215$, but these were measured in neon with a very peculiar geometry.

A step in the direction of modifying Hull's cathode to make it more useful in electron optical devices is described in French Patent No. 903976, based on a Siemens–Halske application of 1942.*

Here the front of the cathode consists of a large number of tungsten or molybdenum strips (Fig. 9). Behind these is a reservoir of active material which might be an alkaline-earth oxide, a mixture of two or more such oxides, alkaline-earth metals or alloys of such metals. This cathode is distinguished from the laminated cathode previously described by the presence of the reservoir and by the fact that the strips are placed closely together so as to reduce field penetration between them to a negligible amount. The use of porous discs is also mentioned.

A curious point is that the use of nickel is mentioned as an inhibitor of emission, as it is said that barium does not diffuse on nickel. Thus, a shaped emitter is claimed in which the emitting area is molybdenum and the rest nickel.

The first published account of this work states that the work function of barium on molybdenum was found to be 1·4 eV against 1·7 eV for barium on tungsten (Katz[33]). This paper also contains results on a cathode made by overwinding tungsten wire on a tungsten mandrel sprayed with thoria.

* I am deeply indebted to Dr. H. Katz of Siemens and Halske for information on his early work in the field, which seems fully to anticipate the better-known work of Lemmens.

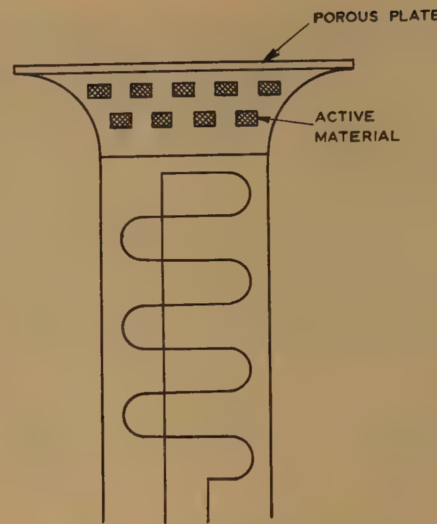


Fig. 9.—Katz's cathode of 1942.

Independently, the workers in Holland had produced the L-cathode,[12] and it is convenient to use the date of publication of their paper as the beginning of the modern era in high-density cathodes.

Further independent work had also been carried out in the United States. This has already been mentioned as stemming from the desire to improve oxide cathodes for magnetrons. This work did, however, lead Coomes and Forsbergh[34] to try a cathode which is a compromise between the mush cathode and the modern sintered-nickel cathode. These workers mixed 50–60% nickel powder of a rather large particle size (200 mesh) with powdered carbonates and 0·01–0·03% magnesium, which may be an activator. The powder was ball-milled and dried before compressing into pellets at 10–15 tons/in^2. The pellets were sintered in dry hydrogen for about an hour during which time the temperature was taken up to 1 200° C. The temperature was held at 1 000° C while the gas cleared. Cylindrical cathodes were made by assembling the pellets on a mandrel before sintering and by machining down to size.

Activation consisted in heating for several hours at 1 000–1 050° C and then at 950° C. D.C. emissions at 950° C of 0·2 A/cm^2 were obtained, while pulsed emissions up to 25 A/cm^2 were measured.

The cathode described above would have a very high porosity somewhat offset by the machining processes described, which would seal the surface pores. The rate of arrival of active material would be limited mainly by thermo-chemical considerations and partly by the surface polishing. Thus, though the technique of preparation is broadly similar to that of some modern cathodes, the details of operation are likely to be completely different. It should be emphasized that changes of technique which at first sight seem trivial do, in fact, produce first-order changes in cathode performance. A good instance is the combined influence of pressure and metal particle size in determining the pore size and porosity, which in turn determine whether rate of transport of active material or reaction velocity is predominant in a particular system.

(10) THE L-CATHODE

The term 'L-cathode' is currently used to describe dispenser type cathodes which utilize the emission from the barium tungsten system. Several modifications of the cathode are known, so we describe the cathode investigated by Lemmens *et al.* under the title of 'original L-cathode'.

(10.1) Original L-Cathode

The structure is shown in Fig. 10. A porous tungsten disc, prepared by the techniques of powder metallurgy, is attached in a vapour-tight manner to a molybdenum support sleeve. Behind the tungsten disc is an enclosure containing a charge of $(BaSr)CO_3$ as used in conventional cathodes. Behind this enclosure is an ordinary alumina-insulated heater.

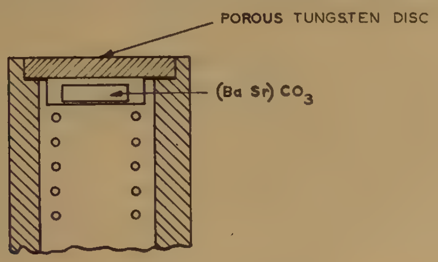


Fig. 10.—L-cathode.

The cathode is prepared by breaking down the carbonate to oxides in the ordinary way, the gas being pumped off through the pores in the tungsten. When the carbonates have been reduced to oxides, chemical reactions between the oxides and the hot tungsten lead to the production of free barium, which flows through the pores of the tungsten disc, eventually reaching the surface. The surface becomes covered with a monolayer of barium and the thermionic emission increases greatly owing to the very marked reduction in work function from about 4·5 eV to 1·5 eV.

The above description is an outline of a process whose details are complicated and sometimes in doubt. We shall return to the details later.

The emission from the original L-cathode is shown in Fig. 1, but the most interesting feature was the power of giving elevated values of continuous emission with reasonable values of life. This feature obviously stemmed from the provision of a relatively large mass of (BaSr)O and a slowing down of the evaporation by the porous plug. Early figures showed that such cathodes give lives of 8 000 h at 1·3 A/cm^2 and a temperature of 950° B.[63]

It was soon found that this cathode had several disadvantages as well as its undoubted advantages. First, it was found to be difficult to attach the tungsten disc to the support in such a manner that big leaks of barium vapour were entirely eliminated. Naturally, if such leaks were present the life would be short. Secondly, it was found that the temperatures in the out-gassing and activating stages were critical. This point was elucidated by the work of Hughes, Coppola and Evans,[35] which showed that a side reaction producing the barium tungstate, Ba_3WO_6, is probable if the temperature for breakdown of carbonate to oxide is raised above 1 010° C. On the other hand, if the temperature is below 1 000° C the breakdown time becomes excessively long. Thus, breakdown must be carried out in a narrow temperature range, 1 000–1 015° C, and the time (about 2 hours) is rather long.

(10.2) Impregnated L-Cathode

Attempts were very soon made to overcome the disadvantages of the L-cathode by incorporating the active material in the body of the cathode. If the mixed carbonates were used, this method failed to produce cathodes with useful emissions; this is now understandable in the light of the considerations given above, as the very intimate and extensive contact between $(BaSr)CO_3$ crystallites and tungsten would favour the production of the tungstate even at temperatures below the range mentioned above. However, Levi[36] found that if the impregnant was a mixture of normal and basic barium aluminates then the emission could be obtained. Further advantages of this cathode, called the impregnated L-cathode, are the much reduced evolution of gas during processing and the simplicity of manufacture (Fig. 11). Levi[37] subsequently found that the addition of calcium to the barium aluminate improved the emission by a factor of about 4 and improved the life. His impregnant had the molar constitution 5 moles BaO, 2 moles Al_2O_3 and 3 moles CaO. Brodie and Jenkins[38] investigated the properties of cathodes using impregnants in which the proportion of calcium oxide was varied, and found that better emissions were obtained from an impregnant in which 0·5 mole of CaO was added to each mole of the basic aluminate 3 BaO . Al_2O_3. They achieved a d.c. emission of about 6 A/cm^2 at 1 340° K. The rate of evaporation of barium was also much less than in the earlier cathodes. Life tests at 1·5 A/cm^2 and 1 200° K have shown no change over 5 000 h. According to Brodie and Jenkins's emission curves, this cathode would be operating at about 50% over the saturation emission. It should be noted in passing that there is no reason why this should not be done with modern dispenser cathodes if the resulting dependence on exact cathode temperature is tolerable. The practice does not seem harmful to the cathode. These authors also tested barium silicates, but the results were inferior to those with aluminates (see also Reference 50 for similar results).

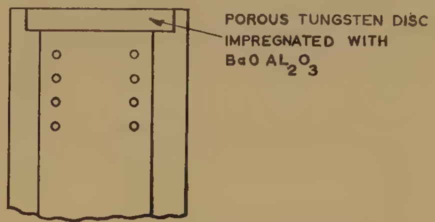


Fig. 11.—Impregnated L-cathode.

The properties of an impregnated L-cathode made by compressing and sintering a powdered mixture of tungsten and active material have been extensively investigated by Coppola and Hughes.[39] We shall quote some of their results later.

So far, we have said nothing about the techniques used for impregnation. These have not been fully discussed in the literature. One technique which has been mentioned is impregnation *in vacuo* in which the cathodes are introduced into a kinetic vacuum system where they are dipped into a bath of the molten impregnant for a time just sufficient to fill the pores. Another technique is to slide a cup containing the powdered components of the impregnant over the front of the cathode. The assembly is then evacuated and the powders are melted by an appropriate furnace or by r.f. heating and the impregnant soaks into the cathode, any excess being subsequently machined off. In the work of Brodie and Jenkins[40] impregnation was carried out in a hydrogen atmosphere. The alkaline earths were supplied as carbonates which were carefully broken down to oxides by pre-heating before the thoroughly mixed powders were fused.

We may remark in reference to all the impregnated cathodes that, although the out-gassing time is short, a very prolonged ageing is necessary before the full emission is obtained. Levi[37] states that the cathode without calcium took three days of high-temperature ageing to produce the final activated state. The use of calcium oxide reduces this to 2–3 hours, and if sufficient current can be drawn while on the pump no further ageing is needed. Brodie and Jenkins found that increasing the molecular

proportion of calcium oxide decreased the time, the effect being small when proportions greater than 0·5 mole CaO are in question. A time in excess of 2 hours must, however, still be considered as long.

Another development of the L-cathode which should be mentioned is one in which the (BaSr)CO_3 charge of the original L-cathode is replaced by a pellet of barium-calcium aluminate with tungsten, as used in the impregnated cathode. This was found by Venema and Van den Broek[41] to give the same emission as the normal L-cathode but with easier processing and less evaporation. The following data are of interest. The active material was of the same composition as used by Levi, namely $5BaO + 3CaO + 2Al_2O_3$, and 20% by weight of this was pressed with 80% tungsten powder into a pellet. The pellet was fired in hydrogen at 2 200° K to sinter it and was then used as a filling for the L-cathode. A very long activation time of a few hundred hours was found necessary, but by preheating the cathode in hydrogen at 2 200° K for a few minutes the activation time was shortened to a reasonable value. Apparently the preheated cathodes could be exposed to atmosphere without impairment of this feature. However, the resulting cathodes, while giving useful performance in EC57 triodes, are rather easily poisoned by gas, this being a consequence of the low evaporation rate.

Further work on various modifications to the L-cathode has been carried out in Japan by a group under T. Hashimoto.[42] These investigators studied the influence of reducing agents added to the filling of the original L-cathode. They concluded that better results were obtained if silicon and carbon were added to the charge. Using 1 mole $BaCO_3$ + 1 mole C + 0·8% Si, emissions of 1·1 A/cm² were constant at 1 000° C after 13 500 h. This group[43] have confirmed the work of Brodie and Jenkins and have investigated other impregnating mixtures. They find $3BaO . Al_2O_3 + \frac{1}{2}MgO$ to be very slightly better than the calcium mixture, and $BaO . BeO_2$ used to impregnate a porous plug containing 10% of tungsten carbide quite noticeably better, the pulsed emissions at 1 000° C being

	A/cm²
$3BaO . Al_2CO_3 + \frac{1}{2}CaO$	1·5–2
$3BaO . Al_2CO_3 + \frac{1}{2}MgO$	2–3
$BaO . BeO_2$	7–10

Unfortunately the evaporation rate increases with the emission.

The cathode of Coppola and Hughes[39] differs from those so far mentioned in that the active material is directly introduced into the matrix by compressing and sintering powdered metal and powdered active material. If the carbonates of barium and strontium alone are used, no activation occurs as the reactions are not those which produce free barium. However, the barium aluminate already mentioned can successfully be used. In this case the matrix has a high porosity, about 40%, and the rate of supply of barium to the surface is governed primarily by its rate of chemical production. In initial experiments it was found that the tribarium aluminate, $Ba_3Al_2O_6$, was favourable from the point of view of barium production but that the incorporation of this material into a tungsten matrix gave high evaporation and short life. A molybdenum matrix, on the other hand, gave low evaporation and negligible emission. The evaporation from cathodes employing mixtures of tungsten and molybdenum was measured, and an alloy containing 25% tungsten was chosen. The cathodes were then pressed from a mixture of 90% by weight of this alloy plus 10% of barium aluminate ($5BaO$: $2Al_2O_3$). Later, additions of calcium oxide as recommended by Levi produced further improvements.

The cathodes are pressed into their retaining cylinders at 70 tons/in² and are then sintered in vacuum or hydrogen at 1 750° B.* At 1 130° B the cathode containing no calcium gives a saturated pulsed emission at 1 kV of 5·5 A/cm², while that containing calcium gives 10·5 A/cm². Richardson constants of $A = 2{\cdot}4$ and $\phi = 1{\cdot}7$ eV were obtained for the calcium cathode.

Activation of these cathodes is simple. The diode is pumped in the conventional way and the cathode temperature raised to about 1 200° B in about 5 min. Current is drawn and initially stabilizes after about a minute. However, further ageing after seal-off is necessary for full activation. Calcium aids this process also, reducing the time at 1 130° B from about 70 h to about 5 h.

Clearly, these cathodes have important manufacturing advantages over the variants previously described. The emission obtained is somewhat less than the best values of Brodie and Jenkins,[44] who measured 20 A/cm² pulsed at 1 130° C for cathodes with a similar active mixture but made by impregnation in hydrogen.

(10.3) The Mechanism of Operation of L-Cathodes

The mechanism of the operation of L-cathodes was initially investigated by the inventors,[12] then by Schaeffer and White[45] and in considerably more detail by Rittner, Ahlert and Rutledge.[46] The major questions are as follows:

(*a*) By what reactions is the free barium produced?
(*b*) How does the barium reach the cathode surface?
(*c*) Is the cathode surface barium a monolayer on tungsten or is it more complicated; for example, barium on oxygen on tungsten?
(*d*) How does the active surface build up on the initially clean tungsten?
(*e*) What constitutes the end of life?

The conclusions of Schaeffer and White differ materially from those of Rittner *et al.*, partly through some numerical errors. In view of these and because of the extensive and careful nature of the Rittner investigation we feel justified in using the conclusions reached by the latter authors. In view of the nature of this survey it is not necessary to do more than give the conclusions here.

(*a*) The required reaction is the breakdown of $BaCO_3$ to BaO:

$$BaCO_3 = BaO + CO_2 \quad . \quad . \quad . \quad . \quad (5)$$

The unwanted side reaction,

$$BaCO_3 + \tfrac{1}{3}W = \tfrac{1}{3}Ba_3WO_6 + CO \quad . \quad . \quad . \quad (6)$$

must be minimized during the outgassing. The second stage is the reduction of BaO by the following reaction:

$$2BaO + \tfrac{1}{3}W = \tfrac{1}{3}Ba_3WO_6 + Ba \quad . \quad . \quad . \quad (7)$$

This reaction eventually goes to completion and this occurrence substantially denotes the end of life because the final possible reaction,

$$\tfrac{2}{3}Ba_3WO_6 + \tfrac{1}{3}W = BaWO_4 + Ba \quad . \quad . \quad . \quad (8)$$

liberates $BaWO_4$, which is thought to be sufficiently volatile to be liberated into the inter-electrode space, where it is decomposed by electron bombardment and poisons the cathode. Strontium salts, when present, assist by reducing the equilibrium barium vapour pressure, consequently prolonging the life, and also by making the breakdown of carbonates to oxides easier.

(*b*) The barium reaches the surface partly by migration over internal surfaces but mainly by Knudsen flow through the pores. The evaporation rate can thus be slightly influenced by the plug thickness but much more importantly by the average pore diameter.

* It is not clear from the original whether the temperature quoted is really brightness (on the molybdenum sleeve) or true temperature in degrees centigrade. For the one value which is unambiguously stated, brightness temperature is given. [We] have assumed that this is done throughout the paper. For molybdenum 1 130° B = 1 210° C = 1 483° K.

(c) The cathode surface is thought to contain oxygen in the absorbed layer. The most direct evidence for this view is obtained by evaporation of Ba + BaO mixtures on to a clean tungsten wire. The maximum emission is obtained for 100% BaO and is a little greater than the value for an L-cathode. 100% Ba gives an emission which is lower by more than an order of magnitude.

(d) Active material flows through the pores and spreads out over the surface from their ends by diffusion. The diffusion length and lifetime have been measured fairly directly. The results, which are given in Fig. 12, show that, since the diffusion length is much greater than the separation between pores, the surface coverage must be 100%.

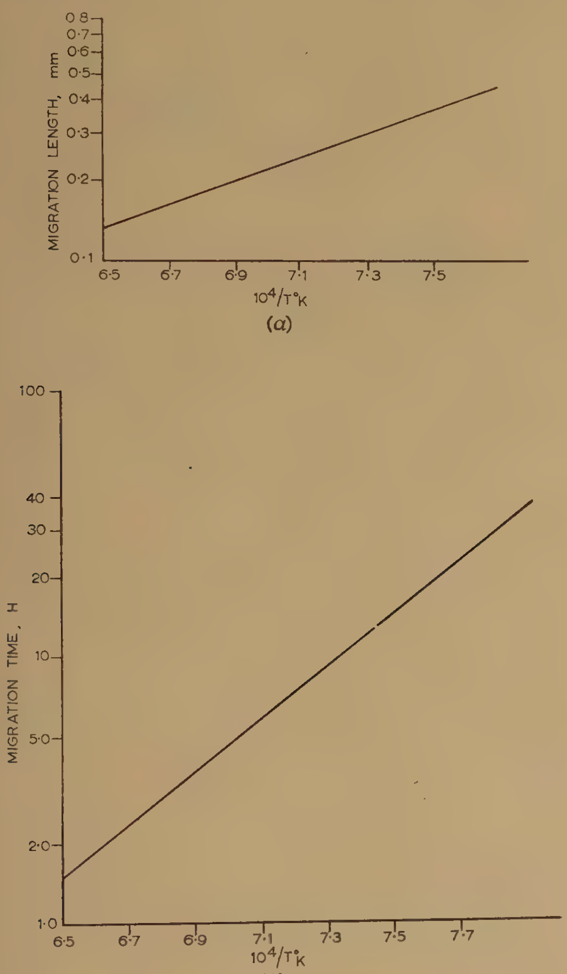


Fig. 12.—Diffusion on L-cathode.

(e) Life ends substantially with the ending of reaction 7.

For the impregnated L-cathode, Brodie and Jenkins believe that the reactions involved are

$$3BaO \,.\, Al_2O_3 = 2BaO + BaO \,.\, Al_2O_3 \quad . \quad . \quad (9)$$

$$6BaO + W = 3Ba + Ba_3WO_6 \quad . \quad . \quad . \quad (10)$$

The presence of calcium oxide is thought merely to lead to an increase in emission from the ends of the pores and a reduction in the evaporation rate.

(11) THE METAL CAPILLARY CATHODE

The metal capillary cathodes of Katz[33, 47, 48] include the L-cathode as a special case. A variety of emitting systems was studied and several different constructional forms were evolved. A typical experimental cathode was made by filling the reservoir of an original L-cathode with thorium metal. This system gave good emissions, but the temperatures required were in excess of 1 400° C. A more interesting variant is that in which a porous tungsten disc is used in conjunction with a filling of substantially metallic barium, i.e. an alloy such as BaAl or BaBe. To avoid excessive reaction rates, heat insulation is inserted between the emitting disc and the reservoir, as shown in Fig. 13. Using the

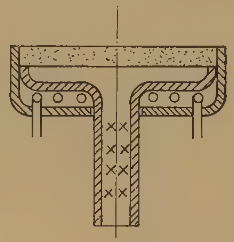


Fig. 13.—Metal capillary cathode.

BaBe alloy, emissions (pulse) of 5 A/cm^2 at 900° C were obtained. Slightly better results were obtained when molybdenum discs were used instead of tungsten. Other work described includes a study of the addition of reducing agents, e.g. silicon, to the active material of the L-cathode.

The emission constants of capillary cathodes have been studied by Benda,[49] who used an interesting technique. It is easily shown that in the retarding field regime the current through a diode only depends on the applied voltage and the anode work function. The surface under investigation was therefore used as anode of the test diode and the work function deduced from retarding field plots. This method gives the arithmetic-mean work function for a patchy surface. For the $(BaSr)CO_3$ + Si system and a tungsten disc, $\phi \simeq 2 \cdot 0$, $A = 100$. These values compare with Richardson plot values of $\phi_R = 1 \cdot 7$, $A \simeq 1 \cdot 0 - 3 \cdot 0$.

The porosity and evaporation from cathodes using tungsten discs have also been studied in the course of this work.[50]

Dr. Katz informs me that his latest cathodes are often made as illustrated in Fig. 14. Here, two tungsten discs with a space between them are used. The lower disc acts as a mechanical filter; it can have a fairly high porosity and does not require to be at all uniform. By this means the emitting disc is freed from

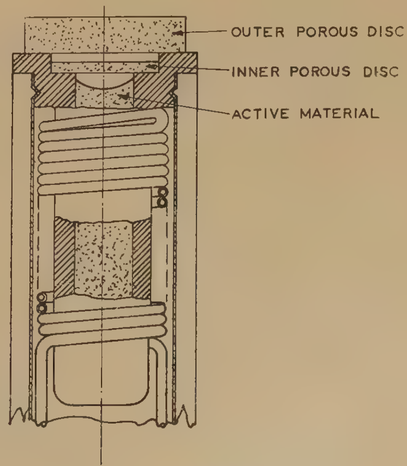


Fig. 14.—Katz's double-disc cathode.

contact with the barium oxide so that no reactions occur and the porosity remains constant. This elaborate construction is, of course, only justified in expensive special-purpose valves, where it gives excellent performance.

(12) DISPENSER CATHODES BASED ON NICKEL

Although cathodes made by compressing active materials with nickel seem to have been investigated by Pirani and Ewest[31] as early as 1928, much less is known about their basic behaviour than about that of similar cathodes based on tungsten. Coomes and Forsbergh[34] investigated cathodes incorporating fairly large amounts of nickel during the war, but did not obtain outstanding results. Also, the mush type of cathode has quite frequently been used in magnetrons, but mush cathodes do not exhibit the characteristic behaviour of dispenser cathodes. The first recent work on dispensers of this type was published by MacNair, Lynch and Hannay.[51] These investigators used a relatively thin layer of nickel matrix on a base of pure nickel (Fig. 15).

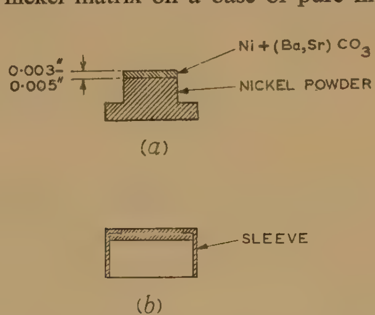


Fig. 15.—Moulded cathode and bariated-nickel cathode.

The layer had to be thin in their case because of the very long time required to break down the active material incorporated in the form of carbonates, and to outgas the system. The reason for this was that carbon was added to the powders to act as a reducing agent. This additive gives a good reduction but leads to excessively long outgassing times in thick cathodes.

Beck, Cutting, Brisbane and King[52, 53] made an extensive study of various metal matrices containing $(BaSr)CO_3$ as active material together with reducing agents. They found that a nickel matrix gave good emission and life when used with suitable reducing agents. Zirconium and titanium hydrides, which had previously been studied as additives in oxide cathodes by Beck and Jones, gave good yields of barium and eliminated the trouble of long outgassing time even with thick cathodes. The resulting cathode was called the b.n., or bariated-nickel, cathode. This has been studied and used by several groups,[54, 55, 56] and we shall discuss it in more detail later.

Balas, Dempsey and Rexer[57] have described a cathode made by forming a porous nickel matrix, usually about 50% porous, impregnating it with a water soluble (BaSr) salt, e.g. the acetate, and precipitating $(BaSr)CO_3$ inside the matrix by adding a water-soluble carbonate, such as ammonium carbonate. With the named salts the reaction by-products are water soluble and decompose at low temperatures. The work function determined from pulsed emission measurements varied from 1·00 to 1·25 eV, and the emission levels were about the same as those found by Beck *et al.* It is again rather doubtful whether, with porosities as high as 50%, one is dealing with a dispenser cathode or with a robust form of oxide cathode. Stout and Gibbons[58] have studied similar cathodes, which they term 'extended interface cathodes'. These authors produce an alloy sponge on a suitable base metal and then fill with oxide. They report that with $(BaSr)O$ the most effective sponge alloy is 0–15% Ti, 100–85% Ni by weight). With ThO_2, 0–50% Zr, 100%–50% W was effective. Their pulsed emissions were similar to those already given.

(12.1) Manufacture of the Bariated-Nickel Cathode

The processes described below are those currently used in the company.

The b.n. cathode is made by compressing the prepared powders into the retaining sleeve.

The preparation of the powders is commenced by cleaning the carbonyl-nickel powder of 1–5 μ particle size. This is done by stoving the powder in a shallow nickel boat, under high vacuum for 2–4 hours at 450° C; if higher temperatures are used the powder aggregates. The nickel powder is then mixed with the desired amounts of $(BaSr)CO_3$, in the form of ordinary double carbonate as used for the preparation of cathode spray, and the reducing agent. We normally use compositions, by weight, in the range 70 Ni: 29 $(BaSr)CO_3$: 1 ZrH_2 to 80 Ni: 19 $(BaSr)CO_3$: 1 ZrH_2, although higher percentages of reducing agent have been investigated. We shall say more about the effects of composition later.

The mixture is milled for a short time in a vibrating ball mill so as to break up the nickel aggregates, if any, using amyl acetate to suspend the mixture. The sludge is filtered through a coarse filter and dried. The powder is then ready for use and can be stored for considerable periods without deterioration.

Two techniques are used for the pressing of cathodes. For cathodes less than 0·5 cm in diameter, all that is needed is a sleeve which is placed in a press tool set up as illustrated in Fig. 16. A measured volume of powder is fed into the sleeve and evenly spread. A pressure of about 80 tons/in^2 is applied and the cathode can then be extracted from the retaining plate and is ready for assembly.

This simple method is not suitable for larger cathodes, as the mechanical strength of large discs of b.n. material thus made is not sufficient to withstand the assembly process. Therefore another technique is used. Here, we proceed as above up to the point where pressure is applied, the pressure in this case being restricted to about 10 tons/in^2. The cathode is removed from the retaining ring and sintered in hydrogen for 20 min at 650° C. The sintering process shrinks the material and strengthens it. The cathode is next pressed a second time to the full 80 tons/in^2. It is then ready for use. The mechanical strength is very much greater and the emission is unimpaired. Care must naturally be observed in handling the cathodes after the first pressing and before the sintering process. Cathodes made in this way are called 'pre-sintered' cathodes.

The greater the metal content of a b.n. powder the better are the mechanical properties of the pressed cathode. Thus, when difficult shapes are in question higher proportions of metal are used, but there is a limit set by the electrical properties, as the porosity, evolution and evaporation of barium are all functions of the percentage of metal.

The retaining sleeve must satisfy certain requirements, both mechanical and electrical. Adhesion is important and it is obviously easy to get good adhesion to nickel sleeves. Molybdenum has a low spectral emissivity and evaporation rate, but the different expansion coefficient means that one has to study the mechanical design of the sleeve carefully. Weak retaining sleeves with castellations are one way of overcoming this difficulty.

The sleeve material has some influence on emission. For instance, we have found that Kovar and nickel-iron, which are useful materials in view of their low thermal conductivity, have a measurable deleterious effect on the emission from the b.n. cathode. This fact was obscured in our early work by variations between cathode and cathode, but we now think that the

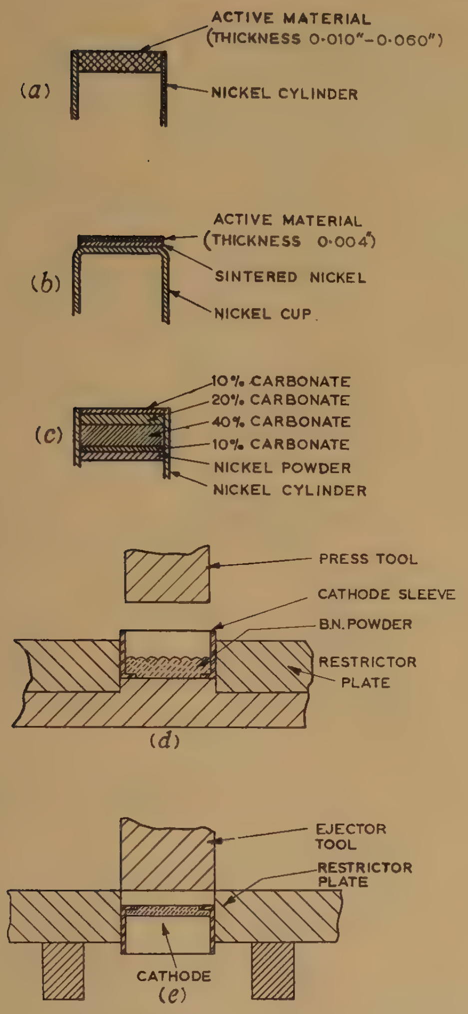


Fig. 16.—Tool arrangement for pressing cathodes, and various cathodes.

(*a*) BN plug 1.
(*b*) BN cathode layer.
(*c*) BN plug 2.
(*d*) Pressing.
(*e*) Ejection.

emission with a Kovar sleeve is only about one-third as great as with a nickel sleeve when both cathodes are fully aged. The use of active nickels such as '0' nickel, is to be deprecated, as the active constituents of the sleeve evaporate out at the higher working temperatures. We therefore now prefer pure nickel for good life and high emission.

It is hardly necessary to stress the importance of cleanliness in all the operations of preparing the cathode. It is especially important to keep the press tools and press platforms free from oil or hydraulic fluid.

The account given above relates to planar cathodes and general principles. We have made concave cathodes for electron guns and a special type of cathode for special electron-optical systems, which we call the 'ridge' cathode, in view of its shape. Fig. 17 illustrates the type of press tool used for making concave cathodes. In this case the end of the sleeve is formed to the required shape before pressing. Care has to be taken to ensure that the pressure is applied to the powder as evenly as possible. Stearic acid has been used as a binder,[18, 55] but while this is undoubtedly effective from the mechanical point of view, it is believed that a distinct lowering of emission is produced. Possibly the stearic acid was insufficiently pure.

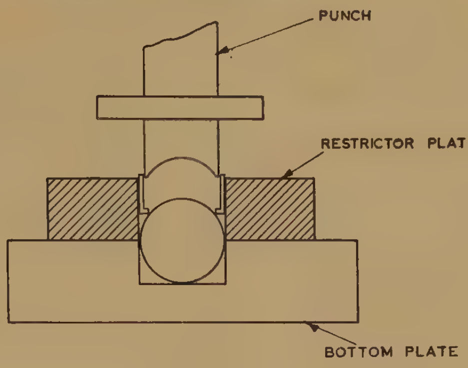


Fig. 17.—Press tools for concave cathodes.

The ridge cathode is illustrated in Fig. 18. This is formed by pressing a pre-sintered cathode against a base-plate into which the desired ridge has been formed by hobbing. Fig. 18 also shows how cathodes of this type are used in precision electron optical systems. A ridge 0·400 in long and of 0·003 in emitting

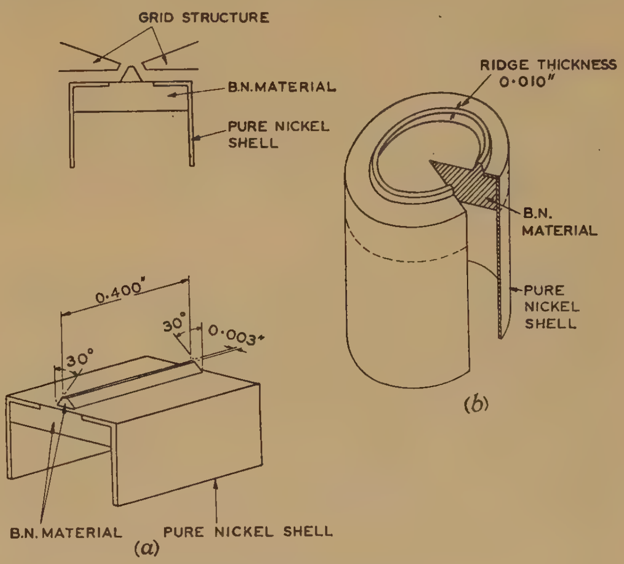


Fig. 18.—Ridge cathodes.

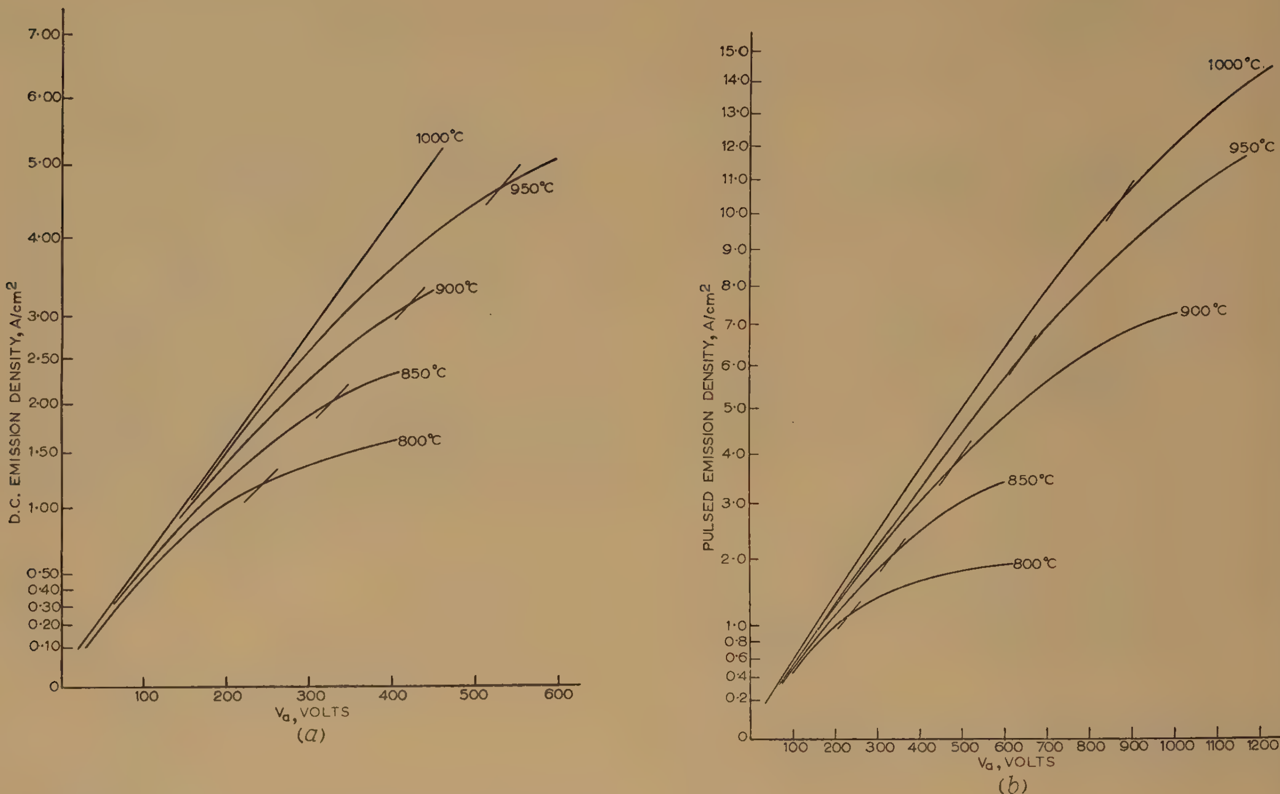
(*a*) Ridge cathode.
(*b*) Annular ridge cathode.

width and 30° slopes has been formed by this method. Such ridges maintain their shape in activation and running and have good lives. Finally, a ridged cathode suitable for the formation of an annular electron beam is shown. Such beams are of considerable interest in modern u.h.f. valves.

(12.2) Processing the B.N. Cathode

B.N. cathodes are easy to process if one remembers that they are far more likely to be damaged by processing at too low rather than too high temperatures and currents. Since outgassing times, etc., depend on the dimensions of the cathode, let us consider the processing of a low-power diode (Fig. 4), in which the b.n. disc is about 0·125 in in diameter and about 0·030 in thick. The outgassing schedule is as follows.

The heater is outgassed by raising the cathode temperature rapidly to about 500° C. The temperature is then raised over a period of 10 min to 1 100° C and maintained at this temperature



until the pressure is reduced to approximately 10^{-5} torr. This takes about 5 min. The anode is next outgassed for 5 min at 1 100° C with the cathode at the same temperature. Then the cathode is heated at 1 200° C for 5 min.

It will be observed that when the cathode is fully outgassed it is flashed to the high temperature of 1 200° C; the reason is interesting but not fully understood. In testing the rate of increase of emission on ageing, it was found that there was an apparently very slow build-up. However, some measurements on diodes with vacuum gauges attached showed that flashing the cathode liberated gas, and after this had cleared the emission was low and another long build-up period was required to reach the initial value. If the flashing process was done on the pump while the anode was maintained at around 1 000° C by eddy-current heating, the diode was fully active immediately on switching on. We concluded that some poisoning agent was liberated by the cathode itself, condensed on the anode and had to be removed by electron bombardment before the emission recovered. Thus the activation build-up was a cleaning of the anode by electron bombardment. With this hypothesis in mind we made tests on sintered-nickel discs without $(BaSr)CO_3$ and found that a visible contaminating layer did, in fact, form on the anode when such discs were flashed. These layers were chemically examined and contained only nickel, so that we were forced to conclude that the only possible poison was oxygen released from one of the more refractory nickel oxides which does not decompose below 1 200° C. This view is compatible with the observation that subsequent 1 200° C flashes are without effect. There seems to be a slight chance that (BaSr)O is removed from those pores which penetrate the surface when the cathode is flashed and that the dissociation of these materials under electron bombardment causes the effect. The evidence against this view is that '10 volt' effect measurements show no critical points at the dissociation potentials of BaO or SrO[59].

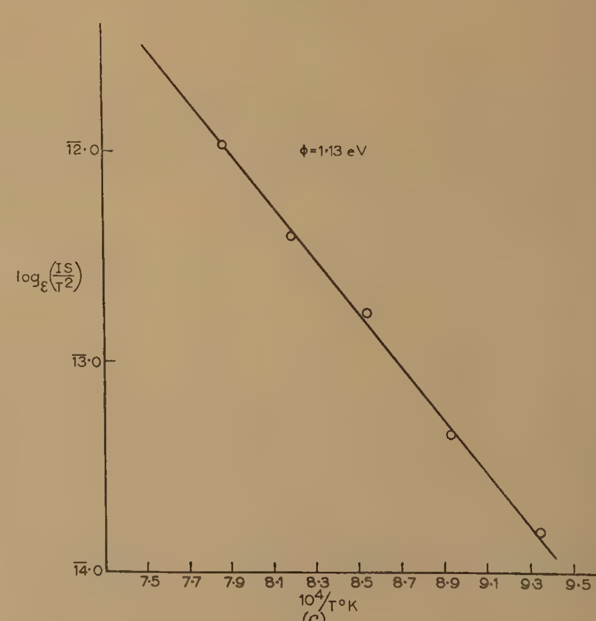


Fig. 19.—Bariated-nickel cathode emission and Richardson plot.

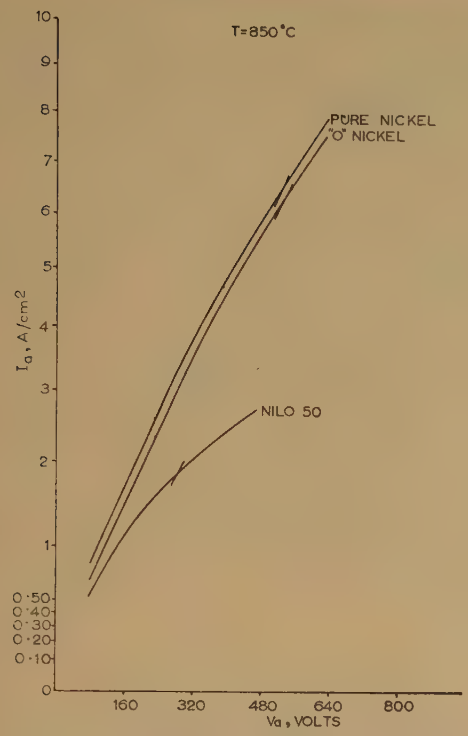


Fig. 20.—Emission for different sleeve materials.

The effect just described is naturally much more prominent in a close-spaced test diode than it is in a valve, especially when the latter uses a focused beam. However, even in this case we believe it wise to break down the cathode with all the neighbouring electrodes at a high temperature.

B.N. cathodes are remarkably insensitive to damage by ion bombardment or by arcing, always provided that the latter does not last long enough to cause local melting. It is thus possible, and often desirable, to use the b.n. cathode for bombarding other electrodes even when the vacuum is not especially good. Moreover, we have used these cathodes in demountable systems and find that the initial values of emission are maintained even after repeated exposures to the atmosphere. However, they have one defect which is not understood, namely a high sensitivity to poisoning by overheating during the sealing-in process. This effect was first noticed when making 17 in cathode-ray tubes using b.n. guns. In the diode gun then under investigation the cathode-grid structure was mounted very close to a hard-glass press using rather thick tungsten supports so that the thermal contact was relatively good.

Great difficulty was experienced in activating the first guns tried. It was then noticed that the stainless-steel grids were discoloured. A dry nitrogen atmosphere was substituted for tank nitrogen and this eliminated the grid discoloration but only partially improved the activation. For a complete cure it was necessary to lengthen the mount and to insert mica shields. These requirements happened to have been incorporated in earlier test valves. In spite of a good deal of work it has not yet been possible to prove whether it is the heating *per se* which is harmful, even in a dry inert gas, or whether the products of combustion from the gas flames combined with the high temperature are the source of the difficulty. It seems probable that a layer of oxide formed on the nickel would be removed by subsequent processing. Perhaps the surface carbonate is reduced and the resulting (BaSr)O reacts with harmful combustion products such as sulphides.

(12.3) Emission from B.N. Cathodes

In our work we have employed three techniques for emission measurement. These are:

(*a*) Testing with semi-sinusoidal pulses whose duration at half amplitude is about 5 microsec.

(*b*) Testing with rectangular pulses of much longer duration, from 100 microsec to 3 millisec.

(*c*) Testing under d.c. conditions.

Differences between the results obtained by the first two methods are conventionally said to be due to pulsed emission decay, while other decay processes, in particular the influence of gas produced from the electrode structure when relatively large powers are dissipated, may have to be invoked to explain the results of d.c. tests.

Fig. 19 gives the results of emission measurements on b.n. cathodes, and Richardson plots based on these measurements. It will be seen that, although the d.c. emission is, in general, below the pulse emission, the work function derived from the direct currents is less than the pulse work function. When b.n. cathodes are tested by technique (*b*) we sometimes observe pulsed emission decay and sometimes do not. When decay is observed, the observations fit the following law, which has been proposed by Déjardin, Mesnard and Uzan:[61]

$$\frac{i - i_\infty}{i} = \frac{i_0 - i_\infty}{i_0} \exp(-\alpha t) \quad . \quad . \quad . \quad . \quad (11)$$

This law was deduced for a model semiconductor including mobile impurity centres. On the other hand, if we were dealing with a barium monolayer, or barium oxide monolayer on nickel, the results should agree with the decay law deduced by Sproul:[62]

$$\frac{1}{i}\frac{di}{dt} = \beta \log\left(\frac{i_0}{i}\right) - \left(\frac{\alpha L}{eNc}\right)i \quad . \quad . \quad . \quad (12)$$

in which case the agreement is not so good.

As has been said, the emission from b.n. cathodes is influenced by the sleeve material. To take a concrete example let us compare the behaviour of cathodes made using '0' nickel, a high-purity nickel and nickel-iron (Fig. 20). The '0' nickel sleeves activate quickly even under adverse conditions and give good emissions, but the evaporation of the impurities is much more rapid than that of nickel, and deleterious deposits may be formed when the cathodes are operated for high emissions. The high-purity nickel is slightly more sensitive to activation conditions, but good results are easily obtained and this is the generally preferred material. With nickel-iron and Kovar the emission is reduced even when very elaborate cleaning techniques are applied to the sleeves. Such cathodes improve somewhat with running, but even after 1 000 h the results are not comparable with nickel.

(12.4) Evaporation Rate and Life

The evaporation rate of active material has been measured by the usual technique of measuring the time taken by a flat tungsten tape stretched over the cathode to reach its maximum thermionic emission. This, following de Boer,[63] is considered to be reached when $1{\cdot}03 \times 10^{-14}$ atoms/cm² or $2{\cdot}34 \times 10^{-8}$ g/cm² have been deposited on the tape. The tube used is illustrated in Fig. 21, which is self-explanatory. Fig. 22 shows results for (*a*) 70% Ni, 29% $(BaSr)CO_3$, 1% ZrH_2 made by the single-pressing technique, (*b*) 80% Ni, 19% $(BaSr)CO_3$, 1% ZrH_2 made as above, and (*c*) 70% Ni, 29% $(BaSr)CO_3$, 1% ZrH_2 made by the pre-sintered technique.

Also shown, for reference, are results of Brodie and Jenkins[40] for an impregnated L-cathode of the sort which gave the best emission. It is seen that in the operating range, i.e.

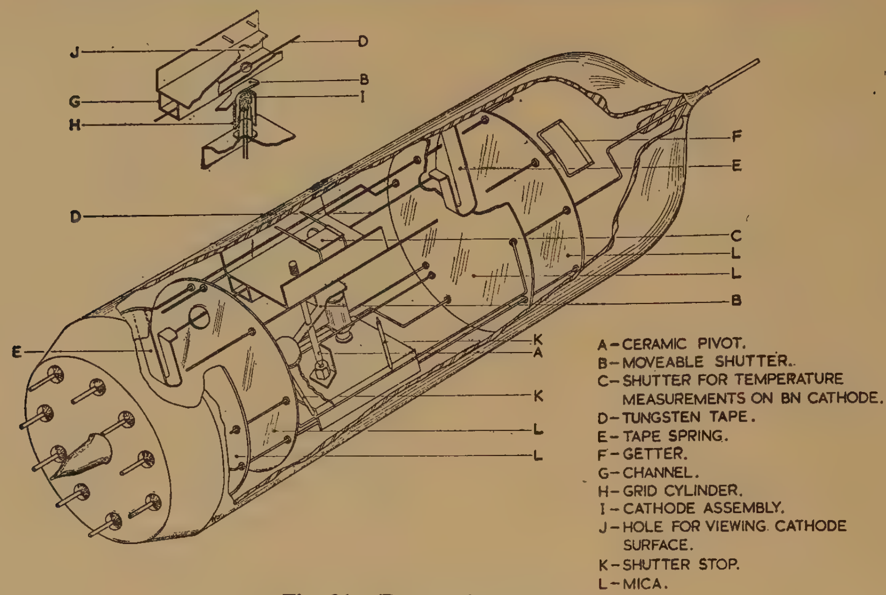


Fig. 21.—Evaporation test valve.

$7\cdot0 < 10^4/T < 9\cdot0$, the b.n. cathodes all have lower evaporation rates and, in the case of the pre-sintered cathode, a much lower evaporation rate, which at 1 250° K is only one-twentieth as great as that of the L-cathode. Still at 1 250° K, we observe that b.n. cathodes take between 3 and 15 hours to evaporate a monolayer of barium. It does not, therefore, seem unreasonable to expect to observe activation and decay processes which last for the same length of time.

Clearly, knowing the evaporation rate we can estimate the time taken to remove all the barium. At 1 250° K for a normal cathode of about 1 mm thickness the figure ranges from 2×10^6 h for curve (*a*) to $2\cdot7 \times 10^7$ h for type (*c*). While these figures have some bearing on the life of the cathode, we must not pretend that we know that such lives can be achieved. The real life depends on the heater life, the rate of sintering, which might for instance close the pores to such an extent that insufficient active material reached the surface, and the possible presence of deleterious chemical reactions which may produce poisons in a manner similar to that indicated by eqn. (8).

Other causes of failure may also be found as our knowledge improves. As we have said, heater life is the most serious problem for continuous emitters. We have found the following lives:

Current Density A/cm²	*Temperature* deg C	*Life* hours	*Failure*
5·0	1 050–1 080	5 000	Heaters
1·0–1·5	920–950	22 000	Heaters
0·3–0·5	800–820	31 000	No failure

Chemical analyses were made of the cathodes, which had run for 22 000 h. While the accuracy was not high, no cathode had lost more than 5% of its initial barium content, yielding a partial confirmation of the evaporation rates mentioned above. It may be added that improved heater designs have since been incorporated and we now hope considerably to exceed 20 000 h at $1\cdot0$ A/cm². We have observed no difficulties due to reaction between the heater and cathode materials; the failure is simply mechanical, induced by long life at high temperature.

(12.5) A Model for the B.N. Cathode

At the present time a model for this cathode must be very tentative. One that best satisfies our present knowledge is the following. The ends of the pores comprise between one-thi[rd] and one-fifth of the emitting area, and an emission which [is] characteristic of an oxide cathode is observed from the pore end[s]. The nickel area between the pores is covered with barium to [an] extent which depends very considerably on the immediate pr[ior] history of the cathode. If a fully covered cathode is pulsed[,] high emission is observed with no decay. However, if the ano[de] is gassy or the diode is maltreated, the coverage is reduced,

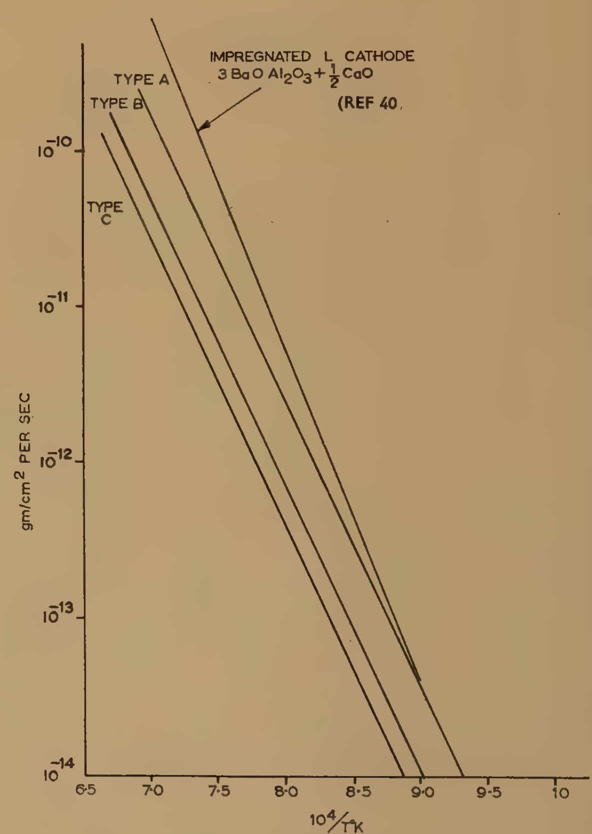


Fig. 22.—Evaporation rates.

the limit only the emission from the pore ends being observed. Decay effects would thus be observed with time-constants which depend on the lifetime of a barium atom on the surface in the set of ambient conditions encountered in the particular case under discussion. The greatest possible decay magnitude would are included in view of their greater reliability. This shows that the emission from the nickel-based cathodes is superior to that from the best tungsten-base cathodes. Our own cathodes do not appear to be as good as those of MacNair *et al.*,[51] but this may be accounted for by differences in measuring technique described

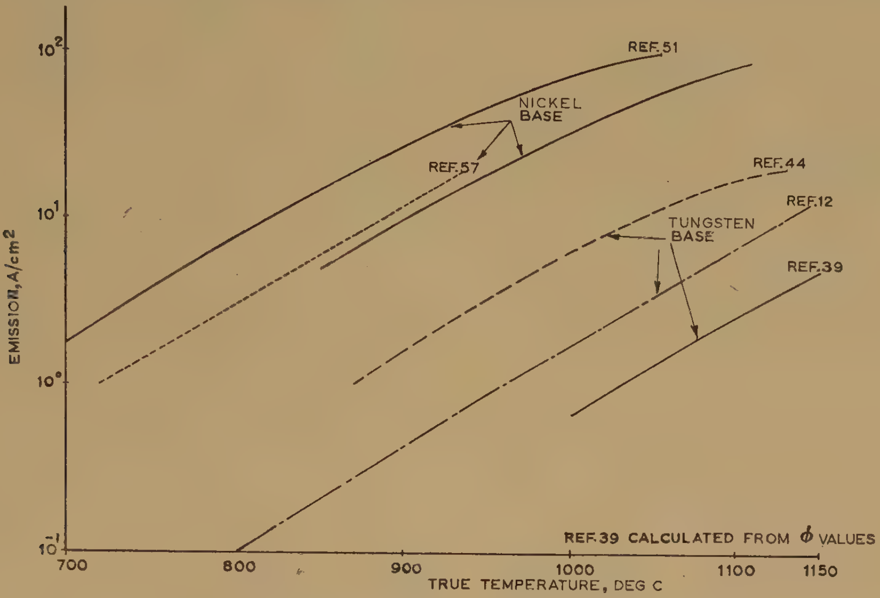


Fig. 23.—Emission for tungsten- and nickel-base cathodes.

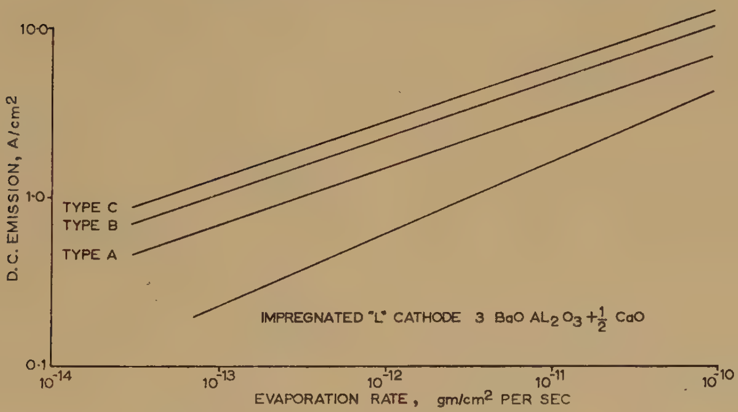


Fig. 24.—Emission versus evaporation rate.

be of the order of the area ratio, which agrees with observation. We prefer to leave open the question whether the emitting surface is barium on nickel or barium on oxidized nickel, but we have no doubt that under some abnormal conditions, such as reactivation after exposure to the atmosphere, one is dealing with emission from thin oxide layers. We are, moreover, in doubt as to the mechanism by which b.n. cathodes appear to be very sensitive to oxygen poisoning before activation and yet will stand pressures of at least 10^{-5} torr of oxygen without any emission change after activation, unless it is that one would observe very much higher values of emission if the vacuum were better than it is in ordinary gettered valves.

(13) A COMPARISON BETWEEN CATHODES BASED ON TUNGSTEN AND THOSE BASED ON NICKEL

Fig. 23 shows emission plots similar to those of Fig. 1 for several types of tungsten and nickel cathode. Pulsed values only

in Section 7. Fig. 24 shows a plot of emission versus evaporation rate, similar to that given by Brodie and Jenkins,[44] with data for several types of b.n. cathode included. Since the emission of the b.n. cathode is greater and the evaporation rate is less, the nickel cathodes are much superior in this respect. However, in fairness one must recognize that the evaporation of the base metal is greater for nickel.

Both types of cathode seem to exhibit pulse emission decay in some circumstances. Though it has often been claimed that the L-cathode is free from decay, the work of Brodie and Jenkins already cited proves that this is not so.

Lives observed in practice seem to depend on heater performace in both cases rather than on the details of the cathode operation. Both types will give lives in excess of 20 000 h at 1–2 A/cm^2, and will give commercially useful lives at higher densities.

The choice between the two systems must therefore be made

on secondary considerations such as manufacturing convenience, cost, ease of processing and freedom from unwanted side effects. No doubt the correct choice will not be the same in all circumstances, and it should be emphasized that the straightforward replacement of an oxide cathode by a dispenser cathode is often not satisfactory.

(14) CONCLUSION

To conclude this survey I should like to make it clear that I believe that, in spite of the advances recently made in the field of high-density emitters, there is still very much work to be done before finality is even approached. There are important gaps in our knowledge which are unlikely to be filled unless fundamental research is carried out. However, emissions at least an order of magnitude greater than those of oxide cathodes, for any specified life, have been achieved, and relatively small improvements in emission would enormously increase the life. I hope that this survey, as well as presenting the situation as it exists to-day, will stimulate further research.

(15) ACKNOWLEDGMENTS

It is a great pleasure to be able to express my indebtedness to my colleagues at Standard Telecommunication Laboratories, especially to Mr. A. D. Brisbane, who has shared several years of hope, pleasure and exasperation while working on cathodes. I also wish to express my thanks to all the many authors in this country and abroad who have assisted me by sending reprints, the names of other workers in the field and notes on unpublished work. I am especially grateful to Dr. Katz of Siemens-Halske for notes on the early history of his work, and to Dr. Danforth of Bartol for sending me a helpful bibliography.

My thanks are due to the Management of Standard Telecommunication Laboratories and Standard Telephones and Cables for encouraging the publication of this survey.

(16) BIBLIOGRAPHY

(16.1) Works Mentioned in the Text

(1) Wright, D. A.: *Proceedings I.E.E.*, Paper No. 1404 R, November, 1952 (**100**, Part III, p. 125).
(2) Nottingham, W. B., in vol. 21 of 'Handbuch der Physik' (J. Springer, 1956).
(3) Hutson, A. R.: *Physical Review*, 1955, **98**, p. 889.
(4) Shelton, H.: *ibid.*, 1957, **107**, p. 1553.
(5) Beck, A. H. W.: 'Thermionic Valves' (Cambridge University Press, 1953), p. 165.
(6) Reed, E. D.: *Bell System Technical Journal*, 1955, **34**, p. 563.
(7) Ekkers, G. F.: I.R.E. Electron Tube Research Conference, 1956.
(8) Nergaard, L. S.: *RCA Review*, 1952, **13**, p. 464.
(9) McNally, J. O., Metson, G. H., Veazie, E. A., and Holmes, M. F.: *Bell System Technical Journal*, 1957, **36**, p. 163.
(10) Metson, G. H., Wagener, S., Holmes, M. F., and Child, M.R.: *Proceedings I.E.E.*, Paper No. 1137 R, June, 1951 (**99**, Part III, p. 69). Metson, G. H.: *Proceedings I.E.E.*, Paper No. 1790 R, April, 1955 (**102 B**, p. 657).
(11) Lafferty, J. M.: *Journal of Applied Physics*, 1951, **22**, p. 299.
(12) Lemmens, H. J., Jansen, M. J., and Loosjes, R.: *Philips Technical Review*, 1950, **11**, p. 341.
(13) Honig, R. E.: *RCA Review*, 1957, **18**, p. 195.
(14) Danforth, W. E., and Haddad, T. A.: *Journal of the Franklin Institute*, 1950, **230**, p. 135.
(15) Danforth, W. E., in 'Advances in Electronics', vol. 5 (Academic Press, 1953).
(16) Becker, J. A.: *Reviews of Modern Physics*, 1935, **9**, p. 95.
(17) Crowell, C. R.: *Journal of Applied Physics*, 1955, **26**, p. 1353.
(18) Hull, A. W.: *Physical Review*, 1939, **56**, p. 86.
(19) Collins, J. B., in 'Microwave Magnetrons', vol. 6 of M.I.T. Series on Radar (McGraw-Hill, 1948).
(20) Fisk, J. B., Hagstrum, H. D., and Hartman, P. L.: *Bell System Technical Journal*, 1945, **25**, p. 167.
(21) Huber, H., and Charles, D.: *Le Vide*, November, 1954, p. 234.
(22) Brücke, E., and Scherzer, O: Elektronenoptik' (J. Springer, 1934).
(23) Johnson, J. B.: *Physical Review*, 1920, **17**, p. 420.
(24) Espe, W., and Knoll, M.: 'Werkstoffkunde der Hochvakuumtechnil (J. Springer, 1936).
(25) Schröter, F.: German Patent No. 535163, 1929–1931.
(26) Babcock, N. L., Holshouser, D. F., and Von Foerster, H.: *Physic Review*, 1953, **97**, p. 755.
(27) Mueller, G. E.: *Transactions of the Institute of Radio Engineer* 1953, **ED-4**, p. 33.
(28) Anderson, C. E., and McEwan, A. W.: Federal Telecommunicatior Laboratory Technical Memorandum (unpublished).
(29) Sugata, E., and Nakamura, S.: Institute of Electrical Communica tions, Japan, 1956, April and October; 1957, April.
(30) Nienhold, J.: German Patent No. 403209, 1920–24.
(31) Pirani, M., and Ewest, H.: German Patent No. 529392, 1928–31.
(32) Espe, W., and Evers, F.: German Patent No. 567909, 1927–34.
(33) Katz, H.: *Entwicklungsberichte der Siemens Konzern*, 1951, **14**, p. 12
(34) Coomes, E. A., and Forsbergh, G.: U.S. Patent 2543439.
(35) Hughes, R. C., Coppola, P. P., and Evans, H. T.: *Journal of Applie Physics*, 1952, **23**, p. 635.
(36) Levi, R.: *ibid.*, 1953, **24**, p. 233.
(37) Levi, R.: *ibid.*, 1955, **26**, p. 639.
(38) Brodie, I., and Jenkins, R. O.: *ibid.*, 1956, **27**, p. 417.
(39) Coppola, P. P., and Hughes, R. C.: *Proceedings of the Institute c Radio Engineers*, 1956, **44**, p. 351.
(40) Brodie, I., and Jenkins, R. O.: *Journal of Electronics*, 1957, **2**, p. 45
(41) Venema, A., and Van Der Broek, C. A. M., at the Washingto Conference of the Electro-Chemical Society, May, 1957.
(42) Hashimoto, T., and Uchida, M.: *Reports of the Electro-Communica tion Laboratories of the Nippon Telegraph and Telephone Co.*, 195 **3**, p. 10.
(43) Hashimoto, T.: Private Communication.
(44) Brodie, I., and Jenkins, R. O.: *British Journal of Applied Physic* 1957, **8**, p. 27.
(45) Schaeffer, D. L., and White, J. E.: *Journal of Applied Physics*, 195 **23**, p. 669.
(46) Rittner, E. S., Rutledge, W. C., and Ahlert, R. H.: *ibid.*, 195 **28**, p. 156.
(47) Katz, H.: *ibid.*, 1953, **24**, p. 597.
(48) Katz, H., and Rau, K. L.: *Frequenz*, 1951, **5**, p. 192.
(49) Benda, H.: *ibid.*, 1953, **7**, p. 1.
(50) Katz, H.: *Le Vide*, November, 1954, p. 279.
(51) MacNair, D., Lynch, R. T., and Hannay, N. B.: *Journal of Applie Physics*, 1953, **24**, p. 1335.
(52) Beck, A. H. W., Cutting, A. B., Brisbane, A. D., and King, G *Nature*, 1954, **174**, p. 1010.
(53) Beck, A. H. W., Cutting, A. B., Brisbane, A. D., and King, G *Le Vide*, November, 1954, p. 302.
(54) Thien-Chi, N. C., and Dussaussoy, P.: *ibid.*, November, 1954, p. 29
(55) Bell, I. J., and Brewer, G. R.: 1955 Symposium of the America Chemical Society.
(56) Hadley, C. P., Rudy, W. G., and Stoechert, A. J.: at the Washingto Conference of the Electro-Chemical Society, May, 1957.
(57) Balas, W., Dempsey, J., and Rexer, E. F.: *Journal of Applied Physic* 1955, **26**, p. 1163.
(58) Stout, V. L., and Gibbons, M. D.: Proceedings of 17th Annu Conference on Physical Electronics at M.I.T., March, 1957.
(59) Wargo, P., and Shepherd, W. G.: *Physical Review*, 1957, **106**, p. 69
(60) Koike, Y., and Shibata, T.: *Reports of the Research Institute c Electrical Communication, Tohoku University*, 1955, **7**, p. 67.
(61) Déjardin, G., Mesnard, G., and Uzan, R.: *Cahiers de Physiqu* 1956, **10**, p. 1.
(62) Sproul, R. L.: *Physical Review*, 1945, **67**, p. 166.
(63) Colloquium Report, *Nature*, 1954, **174**, p. 1176.

(16.2) Other Works

(16.2.1) Ceramic Cathodes.

(64) Nelson, D. E.: *Proceedings of the National Electronics Conferenc* 1954, **9**, p. 463.
(65) Cronin, R. J.: *American Ceramic Society Bulletin*, 1951, **30**, p. 234.
(66) Cronin, R. J.: *Le Vide*, 1955, **10**, p. 355.
(67) Danforth, W. E.: *Physical Review*, 1952, **86**, p. 416.
(68) Fan, H. Y.: *Journal of Applied Physics*, 1949, **20**, p. 682.
(69) Lea-Wilson, C. P.: *Journal of Electronics*, 1956, **1**, p. 529.
(70) Rumyartsev, A. P.: *Zhurnal Teknicheskoĭ Fiziki*, 1955, **25**, p. 1898.

(16.2.2) L-Cathodes.

(71) Espersen, G. A.: *Proceedings of the Institute of Radio Engineers*, 195 **40**, p. 248.
(72) Venema, A.: *Le Vide*, 1954, **9**, p. 269.
(73) Rittner, E. S., Rutledge, W. C., and Ahlert, R. H.: *Journal Applied Physics*, 1956, **28**, p. 1468.
(74) Rittner, E. S., and Ahlert, R. H.: *ibid.*, 1958, **29**, p. 61.
(75) Rutledge, W. C., Milch, A., and Rittner, E. S.: *ibid.*, 1958, **2** p. 834.

(16.2.3) Nickel Matrix Cathodes.

(76) Fane, R. W.: *British Journal of Applied Physics*, 1958, **9**, p. 149.
(77) Richardson, J. F.: *ibid.*, 1957, **8**, p. 361.

DISCUSSION BEFORE THE RADIO AND TELECOMMUNICATION SECTION, 19TH JANUARY, 1959

Mr. W. E. Willshaw: I should like to indicate the present stage of usability of the calcium-impregnated tungsten cathode. At 1 A/cm² we are achieving lives of exceeding 5 000 h in operating devices, performance at this life being unchanged from the initial state, and tests being terminated for reasons quite independent of cathode life. Indications are that many thousands of hours are possible, and this type of cathode has led to real advances in practical possibilities in microwave devices.

The author comments on the heater-insulation difficulty presented by the use of these higher-temperature cathodes, but it is not always necessary to use insulation. Self-supporting heaters, where possible, may present other valve problems, e.g. in the provision of heavy-current supply leads, but in some situations will allow the major problem to be avoided. There are, of course, devices in which electron bombardment of the cathode may provide the necessary heating and may limit the life; in these, high-temperature operation is an advantage, and the heater problem is present only for a limited part of the operating time.

One important problem created by the use of higher cathode operating temperatures in some valves is, of course, that of the effects of evaporation of material from the cathode on to surrounding electrodes. This leads to a new set of difficulties which need to be overcome before satisfactory devices of good life can be produced.

The author has not emphasized particularly the improved breakdown performance experienced when these higher-performance cathodes are used. It is perhaps fortunate that the steps taken in the use of more metallic and refractory surfaces are those necessary to improve voltage-breakdown characteristics, since the requirements for these generally become more severe as the current density increases.

Section 8 suggests that primary emission from a magnetron cathode is of no particular significance in operation, but investigation of the more refined aspects of magnetron operation shows that it is of considerable significance, particularly in the build-up phase, and the types of cathode described here are of great importance.

Mr. E. G. Rowe: The practical application of these cathodes, particularly the bariated-nickel ones, substantiates the author's view that, to obtain the full benefit from them, it is necessary to pay particular attention to the design of the heater, the cathode-support structure and associated components in that general region.

We have found it difficult to elucidate any clearly recommended operating temperatures, but it is at least established that the bariated-nickel cathode requires a higher operating temperature than the normal oxide-coated cathode to achieve the advantages which this particular type offers. This, plus the higher thermal emissivity, makes it difficult to get a heater operating temperature within desirable limits as established in normal heater cathode practice. Nevertheless, the difficulties are less than those encountered with dispenser cathodes based on tungsten matrices. In addition, the nickel base introduces higher evaporation rates, which could cause inter-electrode insulation breakdown when using high-voltage electron guns; but it has been possible, by proper design of the mechanical structure, including suitable positioning of the insulators and the provision of evaporation shields, to achieve satisfactory valve life. Therefore, our philosophy is that the necessity to attend to these points represents the cost of the advantage offered by this cathode.

Dr. R. O. Jenkins: We experienced trouble with the original bariated-nickel cathodes in close-clearance valves. This was due to distortion caused by shrinkage of the matrix during decomposition of the carbonates. It was overcome by using a 0·01 in layer of matrix on top of a 0·03 in layer of nickel keyed into holes drilled round the top of the supporting cylinder. In addition, we found that the original cathodes were also variable in activation, owing to oxidation of the reducing agent in the matrix by the carbon dioxide evolved during the outgassing. This variability was considerably reduced by putting the activator in the nickel backing, where it was protected and could subsequently diffuse into the matrix.

Fig. 24 indicates that the bariated-nickel cathode is superior at all temperatures and emissions, but the curve plotted for it completely ignores the nickel which also evaporates with the active material. Apart from adding to the total quantity of material, the presence of the nickel may well have invalidated the estimation of the active material itself, since De Boer's figures for a pure barium layer were assumed. Our calculations indicate that, if the total evaporation rate, including the nickel, is plotted against the emission of the bariated-nickel cathode, the curve intersects that for the impregnated cathode at about 2 A/cm². In other words, allowing a safety margin, the impregnated cathode is better than the bariated-nickel cathode from the aspect of evaporated films at operating emissions above about 1 A/cm².

The importance of the evaporated nickel was confirmed in a disc-seal triode using a bariated-nickel cathode operating at 850° C and with a peak emission of 2 A/cm² and a mean emission of $\frac{1}{3}$ A/cm². It was found that after about 1 000 hours a conducting film deposited from the cathode on to the glass between the grid and anode seals damped the r.f. cavity and reduced the efficiency to a very low value, although the emission was unimpaired. On analysis, the film proved to be mainly of nickel, and this experience emphasizes the need for adequate shielding mentioned by Mr. Rowe.

Mr. W. G. Trodden: Work on the poisoning of tungsten cathodes impregnated with barium calcium aluminate showed that the troublesome common gases are oxygen, carbon dioxide, water vapour and air. With these gases there is a well-defined threshold pressure below which the cathode emission is quite stable and above which poisoning sets in very rapidly with small increments in pressure. With the cathode operating at a true temperature of 1 100° C, oxygen is the most serious poisoner, and at this temperature, drawing a current of 0·3 A/cm², the threshold pressure is 7×10^{-7} mm Hg; water vapour seems to be the next most serious, the threshold pressure under the same conditions being 3×10^{-6} mm Hg; carbon dioxide will begin to poison at 8×10^{-6} mm Hg, and air at about 3×10^{-5} mm Hg.

If one heats the cathode without drawing space-charge current, the poisoning begins at somewhat lower pressures. We found that with hydrogen, nitrogen and carbon monoxide there was no observable poisoning at pressures up to 10^{-3} mm Hg.

After poisoning by air, oxygen and carbon dioxide, the cathodes could be fully re-activated by pumping off the gas in the system and activating in the usual way. However, after water-vapour poisoning we found it was necessary to bake the vessel at the usual temperature of about 500° C to drive off any residual water vapour; having done this, re-activation was quite easy. We also found that a fully activated cathode could be exposed to the atmosphere when cold for quite long periods and then re-activated, provided that the valve had been baked, and this could be repeated a number of times.

Mr. R. C. McVickers: In Fig. 3 the space is usually less than 1 mm. What accuracy is needed to measure the spacing, d, to make these plots significant.

Mr. C. P. Sandbank: Could some of the effects of evaporation and contamination from bariated-nickel cathodes discussed in

Section 12.2 be accentuated by the fact that these cathodes operate at very high current densities? In closely spaced diodes the anodes also operate at very high current densities, so that any high points on the anodes will become very hot and liberate gas during the early stages of ageing. While some of this gas may have come from the cathode, some may have been in the anode initially. In this case, the effect would only be a much more severe form of the phenomenon noticed with oxide-coated cathodes.

Mr. W. H. Aldous: In Section 2 the author states that there is no evidence that increased current density *per se* has a bad effect on life, but the evidence on this point is far from complete. Has the author any comparative figures for the life of the same type of well-pumped valve when run at the two quoted figures of $0 \cdot 5$ and $0 \cdot 05\,A/cm^2$? It is known that the life at $0 \cdot 5\,A/cm^2$ can be good, but may it not be even better at the lower current density, bearing in mind that some small fraction of the conduction is electrolytic?

Mr. P. F. C. Burke: We have been surprised at the life which can sometimes be obtained from a conventional oxide-coated cathode under rather stringent conditions. For example, a considerable number of travelling-wave tubes running at 3 kV and a cathode current density of $400\,mA/cm^2$ are in use and some have had lives exceeding 25 000 hours. Research to enable this sort of performance to be obtained consistently might be well worth while.

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSION

Mr. A. H. W. Beck (*in reply*): I am interested to hear Mr. Willshaw's figures for the life of tungsten-based cathodes. I agree with him that the self-supporting heater is very useful in conjunction with large cathodes such as are used in high-powered klystrons, but considerations of space render their use difficult at the very highest frequencies. As regards breakdown, not only are the dispenser cathodes less liable to arc under high-voltage pulse conditions but they also recover very quickly when the arc is extinguished.

In reply to Mr. Rowe, the operating temperature of a cathode depends on the emission density required and on the factor of safety allowed by the designer. This latter is purely subjective and different designers will make different estimates of the severity of the conditions affecting the cathode in a given valve. Ideally, therefore, one should set the cathode temperature by life testing over a range of temperatures. Since this process is too lengthy and too costly, one can attempt to establish points on a curve giving temperature versus emission by determining the life in a selected valve structure. The results of an early attempt at this are included in Fig. 4 of Reference 53 and these were our initial recommendations. However, as life testing proceeded, the results tabulated in Section 12.4 of the paper showed that our early line gave excessive temperatures and had too high a slope. I would now suggest the following figures:

Current density A/cm^2	Temperature deg C
$0 \cdot 1$	700–750
$1 \cdot 0$	900–950
$5 \cdot 0$	1 030–1 080

These show that, for emissions less than about $0 \cdot 5\,A/cm^2$, the cathode temperature, but not the cathode power, is substantially the same as that required by an oxide cathode.

We have experienced the difficulties, encountered by Dr. Jenkins, caused by shrinkage and have overcome them by similar means and also by attention to the details of the manufacturing technique. Dr. Jenkins's calculation of evaporation assumes that the rate of evaporation of nickel from a fully active bariated-nickel cathode is that which would be observed from a pure nickel plate of the same area. This assumption appears to be disproved by the following simple experiment. If a bariated-nickel cathode, without any shields, is placed at the centre of an evacuated glass bulb and is run for several thousand hours, it is found that the metallic layer on the glass is much more dense in the regions which have received matter evaporated from the support sleeve than it is in those regions which have received matter evaporated from the emitting surface. This is true for an extremely pure nickel sleeve as well as for 'O' nickel, in which the volatile contaminants make the difference very striking. Our view is, therefore, that Fig. 24 gives a correct evaluation of the behaviour of the emitter proper. The evaporation problems, mentioned by several speakers, are, we believe, due to the supporting members.

In reply to Mr. Trodden, we have made some measurements of the onset of oxygen poisoning and our result, while not of high accuracy, was around 10^{-5} torr, an order of magnitude greater than his result. I am in agreement with his other remarks.

In reply to Mr. McVickers, the design of the diodes is such that the spacing between anode and cathode can be measured with a projection microscope. Mr. Sandbank may be right in suggesting that the anode can contribute to the effects discussed in Section 12.2, but our anode materials were, in fact, processed to a degree of cleanliness far higher than that used in receiver valves.

I agree with the remarks of Mr. Burke on the life of oxide-coated valves at densities approaching the limit, and these and similar observations were the basis of the comment in the paper, which is questioned by Mr. Aldous.

621.382:621.317.61

The Institution of Electrical Engineers
Paper No. 2855 E
July 1959

AN INVESTIGATION OF THE DEPENDENCE OF THE CURRENT GAIN OF A PLANE-ALLOY-JUNCTION TRANSISTOR ON EMITTER CURRENT AND FREQUENCY

By F. J. HYDE, M.Sc., Associate Member.

(*The paper was first received* 28*th March, and in revised form* 22*nd December*, 1958.)

SUMMARY

The complex internal current gain, α_d, of a diffusion-type germanium transistor has been derived from measurements of the external short-circuit current gain, α, at frequencies up to 20 Mc/s and for emitter currents between 15 μA and 3 mA, by taking account of the effects of the emitter and collector depletion-layer capacitances and the ohmic base resistance. The resulting frequency dependence of α_d is that expected from unidimensional diffusion theory. At low emitter currents, the flow of r.f. current in the emitter depletion-layer capacitance causes the cut-off frequency of α to be less than one-third that of α_d.

LIST OF PRINCIPAL SYMBOLS

α, α_d = External and internal short-circuit current gains, respectively, for common-base connection.
$(\beta_d)_{l.f.}$ = Low-frequency value of the common-emitter internal current gain.
C_c = Collector depletion-layer capacitance.
C_e, C_{e0} = Emitter depletion-layer capacitance and its value for zero applied emitter bias.
D_n, D_p = Diffusion constants for electrons and holes, respectively.
$f_\alpha, f_{\alpha d}$ = Cut-off frequencies of the external and internal common-base current gains, respectively.
I_e = Emitter direct current.
k = Boltzmann's constant.
l_n, l_p = Diffusion lengths of electrons and holes, respectively.
n_e = Equilibrium concentration of electrons in the emitter region.
N = Local donor density in the base.
p = Local hole density in the base adjacent to the emitter.
p_b = Thermal-equilibrium hole density in the base adjacent to the emitter.
e = Electronic charge.
$r_{bb'}$ = Ohmic base resistance.
T = Absolute temperature.
τ_n, τ_p = Lifetimes of electrons in the emitter region and holes in the base, respectively.
V_e, V_c = Emitter-to-base and collector-to-base voltages, respectively.
$\phi_\alpha, \phi_{\alpha d}$ = Phase angles of α and α_d at f_α and $f_{\alpha d}$, respectively.
w_0 = Active base width.
w = Depth of penetration of the collector depletion-layer into the base.
$\omega = 2\pi f$ = Angular frequency.
$x = \omega w_0^2/2D_p$.
$x_{\alpha d} = \omega_{\alpha d} w_0^2/2D_p$, where $\omega_{\alpha d} = 2\pi f_{\alpha d}$.
y_i, y_r, y_f, y_o = Common-base parameters of the internal transistor.
Y_n = Internal emitter input admittance arising from the flow of emitter electron current.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.
Mr. Hyde is now in the Department of Electronic Engineering, University College of North Wales.

(1) INTRODUCTION

The complex internal current gain, α_d ($\equiv [-i_o/i_i]_{v_o=0}$), of a *p–n–p* germanium alloy-junction transistor of the diffusion (homogeneous base) type[1] has been determined from measurements of the external near-short-circuit current gain, α ($\equiv [-i_o'/i_i']v_o' \simeq 0$), over a range of emitter current from 15 μA to 3 mA and for frequencies up to 20 Mc/s. The currents i_o, i_i, i_o' and i_i' and the voltages v_o and v_o' are defined in Fig. 1—

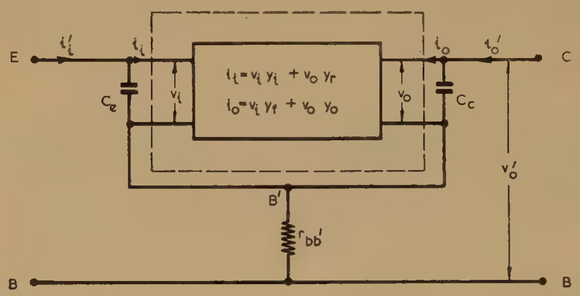


Fig. 1.—Small-signal representation of the transistor.
Internal transistor is within dashed lines.

a schematic of the transistor. Measurement of α was effected by comparing, in amplitude and phase, the r.f. voltages developed across two equal small resistances R in the collector and emitter circuits.[2] The subsequent determination of α_d from α involves consideration of the effects of the external transistor parameters C_e, C_c and $r_{bb'}$ and of R. The effects of C_c, $r_{bb'}$ and R have already been considered, in relation to certain commercial transistors, in an earlier paper.[2] In the present work the role played by C_e—a frequently overlooked parameter in transistor equivalent circuits—is emphasized.

The transistor was designed to have alloy junctions which were plane and parallel to the (111) crystal plane of the base material, whose resistivity was approximately 2 ohm-cm. Furthermore, it was made symmetrical and had a large junction-diameter to active-base-width ratio ($\simeq$ 25), so that the effects of the recombination surface[4, 5, 6] surrounding the emitter would be minimized. It is considered that with this construction a direct comparison between experiment and unidimensional theory should be possible. The comprehensive low-level diffusion theory (i.e. $p/N \ll 1$) was first presented by Early,[7] while Webster[8] and Rittner[9] showed that this theory required modification to include the effect of the electric field which is built up across the base with increasing injection level (p/N no longer very much less than unity). To a close approximation

the low-level formulae of Early[7] may be used at all levels of injection, provided that the diffusion constant D_p of holes in the base is given an effective value $D_p' = D_p(N + 2p)/(N + p)$. In addition, it may be necessary to consider that the lifetime of holes in the base is τ_p', which differs from the low-level value τ_p. Armstrong *et al.*[10] have shown that, for at least some germanium transistors, there is a rise of hole lifetime with increasing inflection-level.

(2) THE THEORETICAL RELATIONSHIP BETWEEN THE INTERNAL AND EXTERNAL SHORT-CIRCUIT CURRENT GAINS

It may be shown[2] that at high frequencies, defined by $\omega\tau_p \gg 1$, α and α_d are related as follows:

$$\alpha \simeq \frac{\dfrac{\alpha_d}{1 + j\omega C_e/y_i} + j\omega C_c r_{bb'}}{1 + j\omega C_c(R + r_{bb'})} \quad . \quad . \quad . \quad (1)$$

The correction of measured values of α for R, $r_{bb'}$ and C_c at each frequency, to yield $\alpha_d/(1 + j\omega C_e/y_i)$, is straightforward. The further correction to determine α_d will now be considered.

(2.1) Consideration of $j\omega C_e/y_i$

With $\omega\tau_p'$ considerably greater than unity, for which condition only is $j\omega C_e/y_i$ significant for r.f. transistors, y_i is given by

$$y_i \simeq G(1 + j)x^{1/2} \coth [(1 + j)x^{1/2}] + Y_n \quad . \quad . \quad (2)$$

where $x = \omega w_0^2/2D_p'$

$$G \equiv A_e(e^2 D_p' p_b/w_0 kT) \exp (eV_e/kT) \simeq eI_e/kT$$

for I_e considerably greater than the reverse saturation current of the emitter junction, provided that Y_n/G and w_0/l_p' are considerably less than unity, and

$$Y_n/G \equiv D_n n_e' w_0(1 + j\omega\tau_n)^{1/2}/D_p' p_b l_n \simeq D_n n_e' w_0/D_p' p_b l_n$$

For large x, the real and imaginary parts of y_i are equal, and increase as $\sqrt{\omega}$.

For an abrupt p–n junction, such as we are considering here, the theoretical expression for C_e is[7,11]

$$C_e = C_{e0}[V_0/(V_0 - V_e)]^{1/2} \quad . \quad . \quad . \quad (3)$$

where C_{e0} is the emitter depletion-layer capacitance for zero emitter bias and V_0 is a constant whose magnitude depends on the resistivity of the base material.

From a consideration of eqns. (2) and (3) it is apparent that $j\omega C_e/y_i$ cannot be ignored compared with unity in the following circumstances: (*a*) at small I_e (small G), (*b*) at very high frequencies (large x), and (*c*) at very large I_e (large C_e). In practice, condition (*c*) is unlikely to arise unless the transistor is operated in a bottomed condition or is pulsed, because, for steady-state operation with normal collector voltages, the maximum permissible collector dissipation severely limits the currents that may be passed.

(3) MEASUREMENTS

(3.1) External Short-Circuit Current Gain

The external short-circuit current gain was measured by the method referred to in Section 1,[2] at frequencies between 1 and 20 Mc/s, for emitter currents between 15 μA and 3 mA, with $R = 50$ ohms. The collector-to-base voltage was maintained at -5 volts, the small voltage drop across the ohmic part of the base resistance being ignored. Measurements were made at room temperature, i.e. 22 ± 2°C.

(3.2) Collector Depletion-Layer Capacitance and Ohmic Base Resistance

The product $C_c r_{bb'}$ was measured using the h-type neutralizing method described by Turner[12] at a frequency of 4 Mc/s. A low emitter current of 100 μA was used to ensure that the collector diffusion capacitance was quite negligible. The value obtained for $C_c r_{bb'}$ was 1850×10^{-12} sec with $V_c = -5$ volts. A value of 97 ohms was derived for $r_{bb'}$ from z-type neutralization of the transistor at 7 Mc/s, using the technique described by Molozzi *et al.*,[13] with $V_c = -1$ volt. The value of C_c follows as 19·1 pF.

(3.3) Emitter Depletion-Layer Capacitance

A direct measurement of the emitter depletion-layer capacitance is not possible with forward emitter bias because a non-zero emitter diffusion capacitance is then associated with it. In these circumstances, however, C_e may be inferred from measurements of the product $C_e r_{bb'}$ (as for $C_c r_{bb'}$ above) with reverse emitter bias, and with the collector functioning as an emitter. Use was made of the theoretical expression for C_e [eqn. (3)], an extrapolation for positive V_e being made. The linear plot of $1/C_e^2$ versus V_e, up to a reverse emitter bias of -2 volts, is

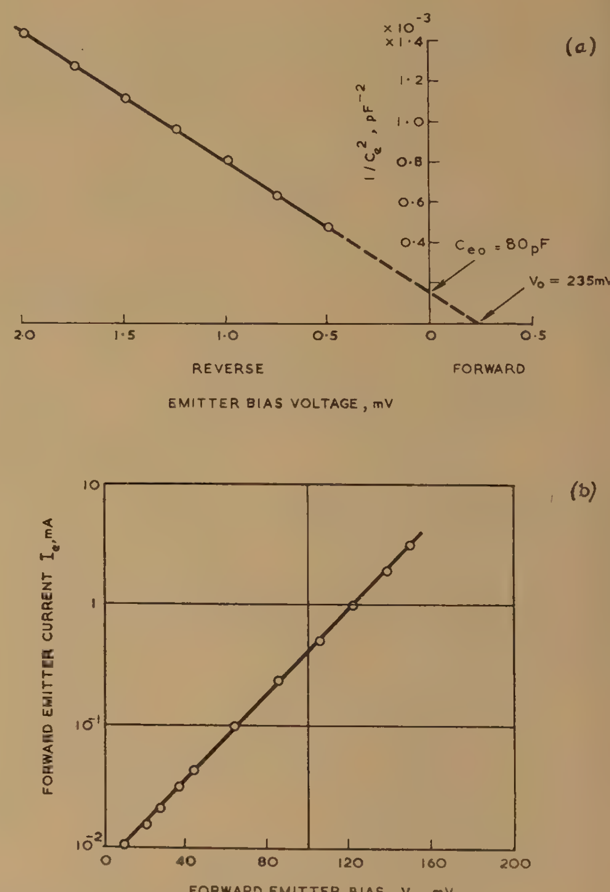


Fig. 2.—Emitter measurements.
(*a*) Dependence of $1/C_e^2$ on emitter voltage, V_e.
(*b*) Static V_e/I_e characteristic of the emitter-to-base junction.

shown in Fig. 2(*a*). From this curve the characteristic parameters may be seen to be $C_{e0} = 80$ pF and $V_0 = 235$ mV. This value of V_0 is that expected[11] for a donor concentration of $8 \times 10^{14}\,\text{cm}^{-3}$ which occurs in transistor-grade n-type germanium of 2 ohm-cm resistivity. The d.c. forward characteristic of the

emitter junction relating V_e to I_e, with $V_c = -5$ volts, is shown in Fig. 2(*b*). Using this in conjunction with Fig. 2(*a*), C_e may be calculated for any value of I_e. Because the transistor is symmetrical, having approximately equal emitter and collector areas, we may note that the value of 19·1 pF for C_c at $V_c = -5$ volts is consistent with $C_{e0} = 80$ pF and $V_0 = 235$ mV given above for the emitter junction.

(3.4) The Internal Emitter Input Admittance

It is not possible to measure y_i directly, although it may be calculated from measurements of any one set of external four-pole parameters at the frequency in question, in conjunction with a knowledge of C_e, C_c and $r_{bb'}$.[14] In the present case it was found expedient to assume the theoretical form of eqn. (2), y_i being determined from this by substituting the values of G, x and Y_n that were deduced from low-frequency measurements, as described in Section 3.5.

(3.5) Internal Base-to-Collector Current Gain at Low Frequency, $(\beta_d)_{l.f.}$

The internal base-to-collector current gain at low frequency was measured for emitter currents between 10 μA and 3 mA with V_c maintained at -5 volts, using a bridge method.[15, 6] The measurement is made in terms of a small fixed resistance, R_2, and variable resistance and capacitance components, R_1 and C. The bridge balance conditions are:[6]

$$\frac{R_2}{R_1} = \mathscr{R}\left[\frac{1}{(\beta_d)_{l.f.}}\right] \simeq \frac{w_0^2}{2D_p'\tau_p'} + \frac{Y_n}{G} \quad . \quad . \quad (4a)$$

$$CR_2 - \frac{C_e}{G} = \frac{1}{\omega}\mathscr{I}\left[\frac{1}{(\beta_d)_{l.f.}}\right] \simeq \frac{w_0^2}{2D_p'} \quad . \quad . \quad (4b)$$

$$CR_1\left(1 - \frac{C_e}{GCR_2}\right) = \frac{1}{\omega}\,\frac{\mathscr{I}\left[\frac{1}{(\beta_d)_{l.f.}}\right]}{\mathscr{R}\left[\frac{1}{(\beta_d)_{l.f.}}\right]}$$

$$\simeq \frac{w_0^2}{2D_p'}\left(\frac{w_0^2}{2D_p'\tau_p'} + \frac{Y_n}{G}\right)^{-1} \simeq \tau_p', \text{ when } \frac{Y_n}{G} < \frac{w_0^2}{2D_p'\tau_p'} \quad (4c)$$

Values of G, which are required to complete the numerical determination of the left-hand sides of eqns. (4*b*) and (4*c*), and for the calculation of y_i, were measured at 1 kc/s, using the bridge method described by Boothroyd and Almond.[16] Over the whole range of I_e from 10 μA to 3 mA, G was found to depend on I_e approximately as $G = I_e/25$ mhos (where I_e is in milliamperes), as was expected on theoretical grounds for low injection levels. The parameter x, required in Section 3.4, was determined using eqn. (4*b*), since

$$x = \omega w_0^2/2D_p' = \omega(CR_2 - C_e/G) \quad . \quad . \quad . \quad (5)$$

From a consideration of eqn. (4*a*), it is shown in Section 5.1 that Y_n may justifiably be neglected in eqn. (2).

(3.6) Thermal Resistance of the Collector Junction

The electronic method described by Loofbourrow and Ollendorf[18] was used to determine the thermal resistance of the collector junction, and a figure of approximately 0·25 deg C/mW deduced.

(3.7) Reverse D.C. Characteristic of the Emitter Junction

The reverse d.c. characteristic of the emitter junction was measured with the collector biased normally at -5 volts. A saturation current of 0·15 μA was observed, leakage current being negligible.

(3.8) Sectioning of the Transistor

At the conclusion of the electrical measurements, the transistor was removed from its can and sectioned. After a final polish with 0–1 μ diamond dust the section surface was etched to bring the *p–n* junctions into relief. This process was repeated several times so that a 3-dimensional picture of the shape of the junctions and their spacings could be built up. It was established that the junctions were plane and parallel and of uniform spacing almost up to their peripheries. The junctions were also found to be approximately circular. The following measurements were obtained:

Emitter diameter	= 28 mils = $7{\cdot}1 \times 10^{-2}$ cm.
Collector diameter	= 29 mils = $7{\cdot}4 \times 10^{-2}$ cm.
Junction spacing	= 1·1 mils = $2{\cdot}8 \times 10^{-3}$ cm.

The depth of penetration w of the collector depletion-layer into the base may be calculated from[11] $w = [2(V_0 - V_c)\epsilon\epsilon_0/eN]^{1/2}$. In the M.K.S. system of units, $\epsilon_0 = 8{\cdot}86 \times 10^{-12}$ F/m and $e = 1{\cdot}6 \times 10^{-19}$ C. ϵ, which is the relative permittivity of the base material, has the value 16 for germanium. For a base material of 2 ohm-cm resistivity, $N = 8 \times 10^{14}\,\text{cm}^{-3}$, while from Fig. 2(*b*), $V_0 = 235$ mV. With $V_c = -5$ volts, the value of w is $3{\cdot}4 \times 10^{-4}$ cm, so that the active base width w_0 is $2{\cdot}5 \times 10^{-3}$ cm.

(4) RESULTS

In Figs. 3(*a*) and (*b*), the full curves show the dependence on emitter current of f_α and ϕ_α—the frequency and corresponding phase angle at which the modulus of α is reduced to $1/\sqrt{2}$ times

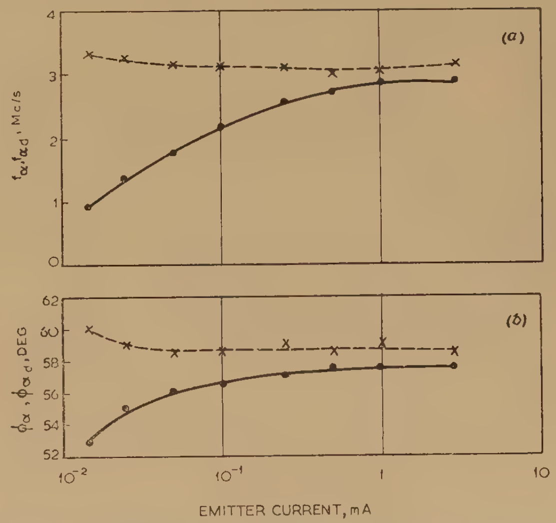
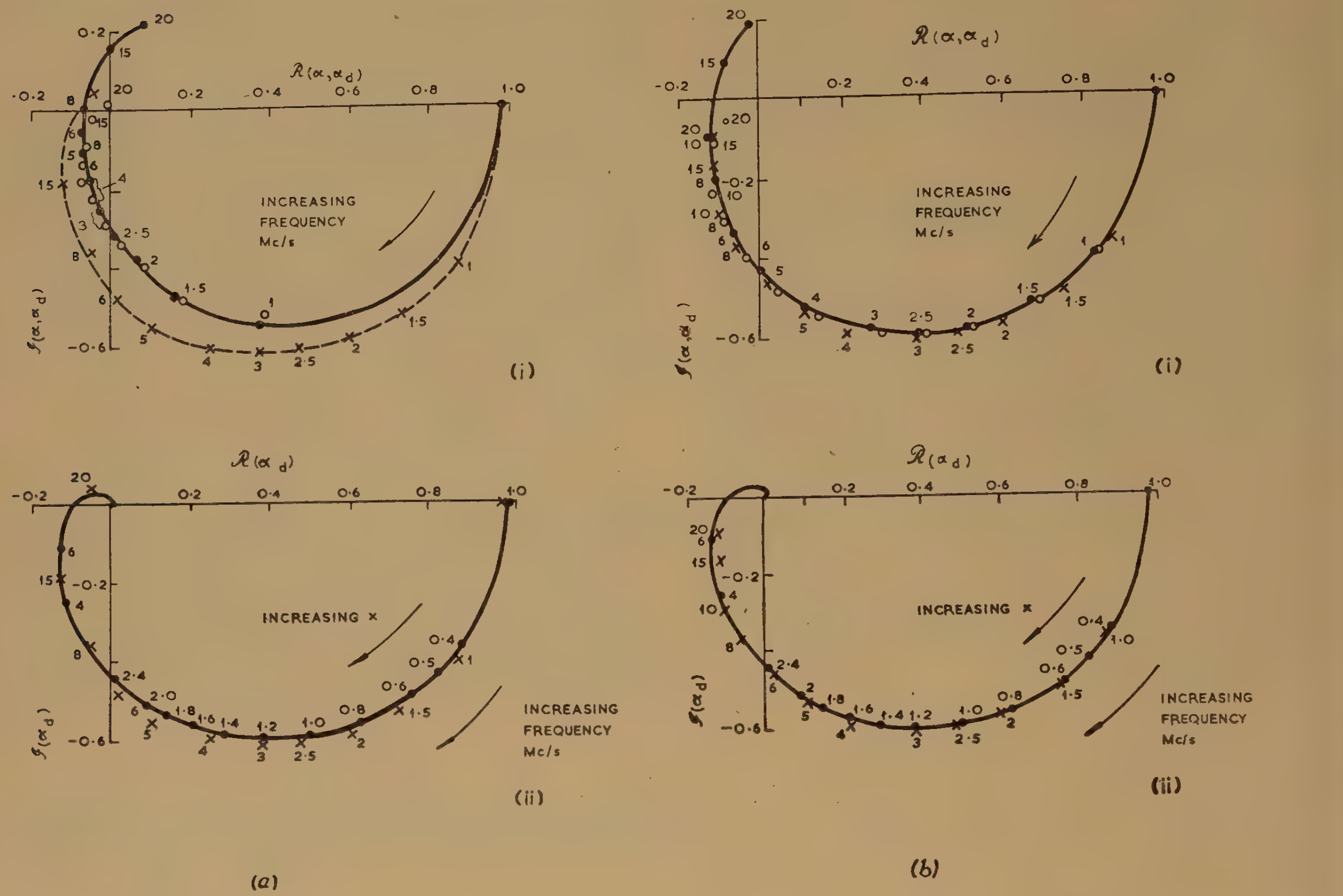


Fig. 3.—Dependence on emitter current of f_α, $f_{\alpha d}$, ϕ_α and $\phi_{\alpha d}$.
(*a*) ——— f_α; — — — $f_{\alpha d}$
(*b*) ——— ϕ_α; — — — $\phi_{\alpha d}$

its zero-frequency value. The individual experimental values (•) of these parameters were derived from complete loci of α covering the frequency range from 5 kc/s to 20 Mc/s. Three of these, for emitter currents of 15 μA, 250 μA and 3 mA, are shown as the full curves in parts (i) of Fig. 4. In the same graphs, $\alpha_d/(1 + j\omega C_e/y_i)$, which is derived from α by taking account of R, $r_{bb'}$ and C_c at each frequency, is shown as circles (○), and α_d, derived from this, is shown as crosses (×). The dashed



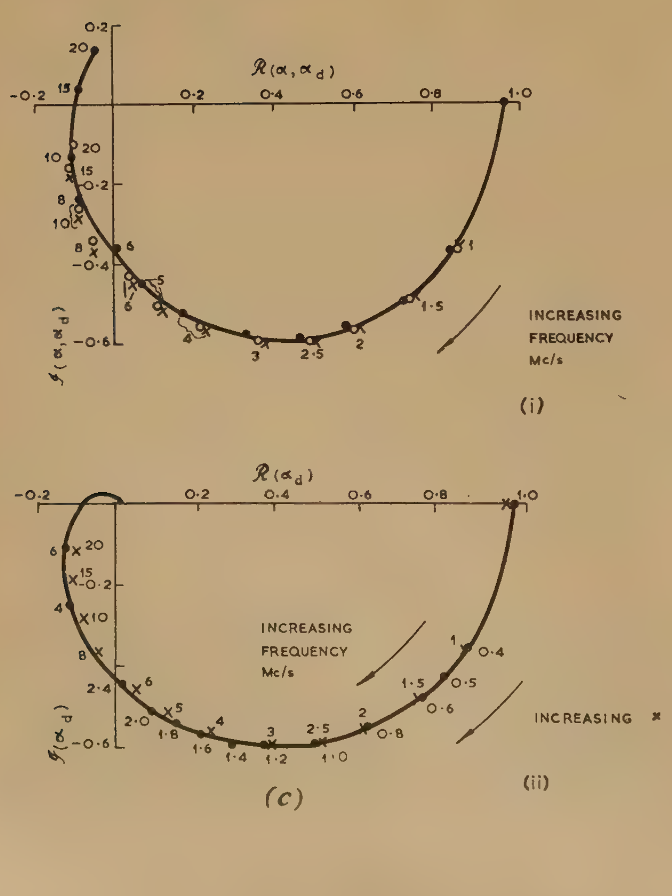


Fig. 4.—Dependence of current gain on frequency.

(i) —●—●—●— External short-circuit current gain α.
○ α corrected for R, $r_{bb'}$ and C_c.
— × — — × — Internal short-circuit current gain α_d.

(ii) Comparison of experimental loci (X) and theoretical limiting form (—●—●—●— of α_d.

Parameters: $V_c = -5$ volts, $T = 22°$ C.
(a) $I_e = 15\,\mu A$; (b) $I_e = 250\,\mu A$; (c) $I_e = 3$ mA.

curves represent the loci of α_d. From these, $f_{\alpha d}$ and $\phi_{\alpha d}$ have been determined and are plotted in Figs. 3(*a*) and (*b*), smooth dashed curves being drawn through them. Here $f_{\alpha d}$ is the frequency at which $|\alpha_d|$ is $1/\sqrt{2}$ times its zero-frequency value, and $\phi_{\alpha d}$ is the corresponding phase-angle. In parts (ii) of Fig. 4, the experimental loci of α_d are repeated and compared with the theoretical curve calculated from[14] $\alpha_a = \alpha_{d0}\,\mathrm{sech}\,\sqrt{(j2x)}$; on this curve the parameter x is shown.

The data obtained from the low-frequency bridge measurements are shown in Fig. 5 as functions of emitter current. In

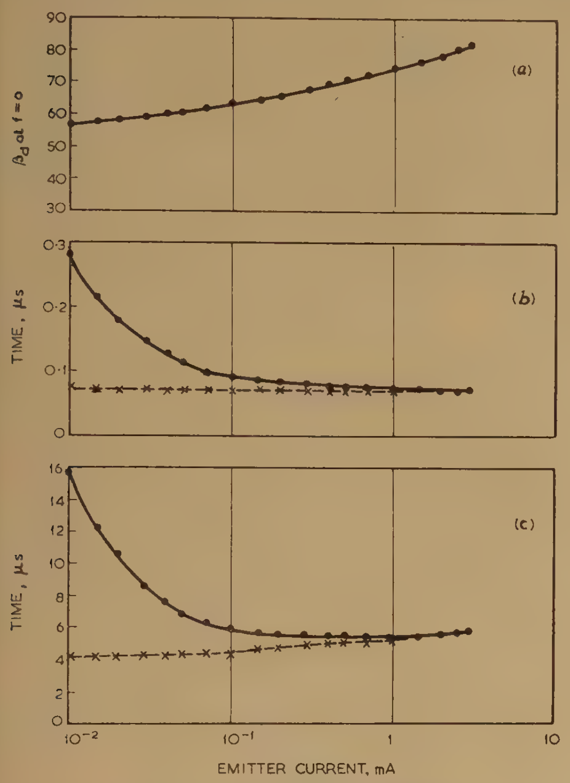


Fig. 5.—Low-frequency bridge data.

(*a*) Variation of $\beta_d|_{f=0}$ with I_e.
(*b*) Variation of CR_2 with I_e.
--- Variation of $CR_2 - C_e/G \simeq w_0^2/2D_p'$ with I_e.
(*c*) Variation of CR_1 with I_e.
--- Variation of $CR_1(1 - C_e/GCR_2) \simeq \tau_p'$ with I_e.

part (*a*) $\mathscr{R}(1/\beta_d)^{-1}_{l.f.} \equiv R_1/R_2$, which is approximately the zero-frequency value of the base-to-collector current gain, is presented: in part (*b*) the full curve shows the variation of CR_2, and the dashed curve that of $CR_2 - C_e/G = (1/\omega)\mathscr{I}(1/\beta_d)_{l.f.} \simeq w_0^2/2D_p'$: in part (*c*) the full curve shows the variation of CR_1, and the dashed curve that of $CR_1(1 - C_e/GCR_2)$, which is approximately τ_p' since it is shown in the next Section that the contribution of Y_n/G to the right-hand side of eqn. (4*c*) can be ignored.

(5) INTERPRETATION OF RESULTS

(5.1) Low-Frequency Bridge and Sectioning Data

The zero-frequency current gain increases monotonically with I_e [Fig. 5(*a*)]. From Figs. 5(*b*) and (*c*) it may be seen that this increase is due to an increase of τ_p' rather than D_p', the latter varying little over the whole current range. Theoretically it is to be expected that D_p' will rise with increasing injection-level (increasing p/N), and that $w_0^2/2D_p'$ will fall in consequence, although the extent of the fall will be abated[12] by the decrease in D_p which accompanies a rise in temperature of the base ($D_p \propto T^{-1\cdot33}$). The value of p/N is given by $(p_b/N)\exp(eV_e/kT)$. The appropriate values[17] for p_b and N are 1×10^{12} and $8 \times 10^{14}\,\mathrm{cm}^{-3}$, respectively, at $T = 295°$C, so that, with $V_e = 149$ mV ($I_e = 3$ mA), $p/N = 0{\cdot}16$. Correspondingly, the value for D_p' is 1·14 times D_p. On the other hand, owing to the temperature rise of the collector junction—nominally 4°C at $I_e = 3$ mA (see Section 3.6)—a fall of D_p to approximately 0·99 times its room-temperature value is expected. The observed decrease of $w_0^2/2D_p'$, which is confined to the current range beyond $I_e = 250\,\mu$A, is from 72×10^{-9} to 67×10^{-9} sec, in reasonable agreement with that expected from the above considerations. The invariance of $w_0^2/2D_p'$ at low emitter currents is good evidence that the numerical corrections to CR_2 for C_e/G are valid. In this connection it may be noted that C_{e0} [$= A_e(\epsilon\epsilon_0 eN/2V_0)^{1/2}$], obtained from the measured values of A_e and V_0, is 78 pF, which agrees well with the figure of 80 pF determined by extrapolation from C_e, measured with reverse emitter bias. The hole lifetime in the base, of approximately 5 microsec [see eqn. (4*c*) and Fig. 5(*c*)] is sufficient to ensure the validity of eqn. (2) at frequencies above 1 Mc/s.

Because the current gain has not reached its peak at $I_e = 3$ mA, it is clear that the term Y_n/G in eqn. (4*a*) is not yet predominant at this current, despite the decrease in $w_0^2/2D_p'\tau_p'$ below its low-current value. It may be inferred, therefore, that Y_n/G must be less than 1/82, which is the value of R_2/R_1 for $I_e = 3$ mA, at all emitter currents up to 3 mA. It is clear that Y_n/G may be ignored in eqn. (2), with negligible loss of accuracy.

(5.2) Current-Gain Data

Figs. 3 and 4 show that the effect of $\omega C_e/y_i$ on the relationship between α and α_d is quite large at small emitter currents, as was predicted in Section 2.1. For $I_e = 15\,\mu$A, $f_{\alpha d}$ is more than three times as large as f_α, although the disparity between $\phi_{\alpha d}$ and ϕ_α is very much less. At emitter currents greater than 1 mA, the effect of $\omega C_e/y_i$ is less than that of C_c, $r_{bb'}$ and R at all frequencies.

In parts (ii) of Fig. 4 it may be seen that there is good agreement between the shapes of the experimentally determined and theoretical loci of α_d. At the smallest emitter current of 15 μA, the experimental curve does lie somewhat outside the theoretical curve, but here the magnitude of the correction for $\omega C_e/y_i$ must be borne in mind. At the other currents shown the loci are coincident. For these two currents, the ratio f/x is shown

Table 1

DEPENDENCE OF RATIO f/x ON FREQUENCY

Frequency	Ratio f/x at $I_e = 250\,\mu$A	Ratio f/x at $I_e = 3$ mA
Mc/s	Mc/s	Mc/s
1·0	2·43	2·44
1·5	2·46	2·46
2·0	2·42	2·44
2·5	2·45	2·55
3·0	2·48	2·52
4·0	2·52	2·58
5·0	2·54	2·63
6·0	2·54	2·66

in Table 1 as a function of frequency, where f and x are corresponding values of frequency and the parameter x on the experimental and theoretical curves, respectively. In each case the ratio increases slightly with frequency, but is reasonably

constant up to 6 Mc/s, the values for $I_e = 3$ mA being somewhat the greater, as is expected in view of the slight decrease in $w_0^2/2D_p'$ [Fig. 5(*b*)] between these currents.

By taking 2·5 as the mean value of $f/x \equiv D_p'/\pi w_0^2$ at $I_e = 250\,\mu$A, and $D_p' = 47\,\text{cm}^2\text{-sec}^{-1}$, $w_0 = 2{\cdot}4 \times 10^{-3}$ cm is obtained. This agrees reasonably well with the value determined by sectioning, namely $2{\cdot}5 \times 10^{-3}$ cm. f/x may also be obtained from the low-frequency bridge data. From eqn. (4*b*) this is given by $1/2\pi(CR_2 - C_e/G)$. For $I_e = 250\,\mu$A [$(CR_2 - C_e/G) = 72 \times 10^{-9}$ sec] $f/x = 2{\cdot}2$ Mc/s, which is somewhat less than the values obtained at radio frequencies. The corresponding value for w_0 is $2{\cdot}6 \times 10^{-3}$ cm.

(6) DISCUSSION OF TRANSISTOR OPERATION

Shockley[19] included a treatment of junction depletion-layer capacitance in his original paper. Subsequently Early,[7] Pritchard[14] and Giacoletto[20] included C_e, C_c and $r_{bb'}$ in equivalent circuits of the transistor. A figure of merit, which is widely used to specify the performance of a transistor as a radio-frequency amplifier, is $f_{\alpha d}/r_{bb'}C_c$.[21] With regard to C_e, however, while many authors acknowledge its existence, it is common to ignore it in practice. The present work shows that for large emitter currents this may be a justifiable assumption, but it is certainly not valid at small emitter currents. Recent work on noise is transistors[22] (in which the effects of external parameters were ignored) has shown that an improvement in r.f. noise figure may be obtained by operating at low emitter currents. It is evident, however, that as I_e is made smaller, an increasing proportion of an r.f. input signal will by-pass the internal transistor by flowing in C_e, so resulting in a deterioration of performance.

(7) CONCLUSIONS

It has been clearly demonstrated that there is a calculable relationship between the internal and external current gains of a plane-junction transistor, whose ratio of junction diameter to active base width is sufficiently great for it to be considered a unidimensional device. In particular it has been shown that at small emitter currents the emitter depletion-layer capacitance exerts a marked influence on this relationship over a wide range of frequency. When the corrections have been applied to α to determine α_d, the latter conforms quite well with theory.

(8) ACKNOWLEDGMENTS

The transistor was kindly provided by Mr. T. R. Scott of Standard Telecommunications Laboratories Ltd.

The work described above was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(9) REFERENCES

(1) LAW, R. R., MUELLER, C. W., PANKOVE, J. I., and ARMSTRONG, L. D.: 'A Developmental *p-n-p* Junction Transistor', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1352.

(2) HYDE, F. J., and SMITH, R. W.: 'An Investigation of the Current Gain of Transistors at Frequencies up to 105 Mc/s', *Proceedings I.E.E.*, Paper No. 2548 R, May, 1958 (**105** B, p. 221).

(3) MUELLER, C. W., and DITRICK, N. H.: 'Uniform Plan Alloy Junctions for Germanium Transistors', fro 'Transistors I' (R.C.A. Laboratories, 1956), p. 121.

(4) MOORE, A. R., and PANKOVE, J. I.: 'The Effects of Junctio Shape and Surface Recombination on Transistor Curre Gain, Part I', *Proceedings of the Institute of Rad Engineers*, 1954, **42**, p. 907.

(5) STRIPP, K. F., and MOORE, A. R.: 'The Effects of Junctio Shape and Surface Recombination on Transistor Curre Gain, Part II', *ibid.*, 1955, **43**, p. 856.

(6) HYDE, F. J.: 'Some Measurements on Commercial Tra sistors and their Relation to Theory', *Proceedings I.E.* Paper No. 2438 R, January, 1958 (**105** B, p. 45).

(7) EARLY, J. M.: 'Design Theory of Junction Transistor *Bell System Technical Journal*, 1953, **32**, p. 1271.

(8) WEBSTER, W. M.: 'On the Variation of Junction Transist Current-Amplification Factor with Emitter Curren *Proceedings of the Institute of Radio Engineers*, 1954, 4 p. 914.

(9) RITTNER, E.: 'Extension of the Theory of the Junctio Transistor', *Physical Review*, 1954, **94**, p. 1161.

(10) ARMSTRONG, L. D., CARLSON, C. L., and BENTIVEGNA, M 'P-N-P Transistors using High-Emitter-Efficiency All Materials', *R.C.A. Review*, 1956, **17**, p. 37.

(11) HARTEN, H. V., KOCH, W., RATH, H. L., and SCHULTZ, W 'Barrier-Layer Capacitance of the Ge–In Alloy Rectifier', *Zeitschrift für Physik*, 1954, **138**, p. 336.

(12) TURNER, R. J.: 'Surface Barrier Transistor Measuremen and Applications', *Tele-Tech*, 1954, **13**, p. 78.

(13) MOLOZZI, A. R., PAGE, D. F., and BOOTHROYD, A. R 'Measurement of High-Frequency Equivalent Circu Parameters of Junction and Surface-Barrier Transistor *Transactions of the Institute of Radio Engineers*, 195 **ED-4**, p. 120.

(14) PRITCHARD, R. L.: 'Frequency-Variation of Junctio Transistor Parameters', *Proceedings of the Institution Radio Engineers*, 1954, **42**, p. 786.

(15) EVANS, D. M.: 'Measurements on Alloy-Type Germaniu Transistors and Their Relation to Theory', *Journal Electronics*, 1956, **1**, p. 461.

(16) BOOTHROYD, A. R., and ALMOND, J.: 'A Bridge f Measuring the A.C. Parameters of Junction Transistor *Proceedings I.E.E.*, Paper No. 1681 M, September, 19 (**101**, Part III, p. 314).

(17) MORIN, F. J., and MAITA, J. P.: 'Conductivity and H Effect in the Intrinsic Range of Germanium', *Physic Review*, 1954, **94**, p. 1525.

(18) LOOFBOURROW, K. E., and OLLENDORF, J.: 'Equipment f Measuring Junction Temperature of an Operati Transistor', from 'Transistors I' (R.C.A. Laboratori 1956), p. 353.

(19) SHOCKLEY, W.: 'The Theory of *p-n* Junctions in Ser conductors and *p-n* Junction Transistors', *Bell Syst Technical Journal*, 1949, **28**, p. 433.

(20) GIACOLETTO, L. J.: 'Study of *p-n-p* Alloy Junction Tra sistors from D.C. through Medium Frequencies', *R.C. Review*, 1954, **15**, p. 506.

(21) PRITCHARD, R. L.: 'High Frequency Power Gain Junction Transistors', *Proceedings of the Institute Radio Engineers*, 1955, **43**, p. 1075.

(22) NIELSEN, E. G.: 'Behaviour of Noise Figure in Juncti Transistors', *ibid.*, 1957, **45**, p. 957.

621.382:621.317.61

The Institution of Electrical Engineers
Paper No. 2856 E
July 1959

AN INVESTIGATION OF THE CURRENT GAIN OF A DRIFT TRANSISTOR AT FREQUENCIES UP TO 105 Mc/s

By F. J. HYDE, M.Sc., Associate Member.

(*The paper was first received* 28*th March, and in revised form* 22*nd December*, 1958.)

SUMMARY

The complex internal short-circuit current gain, α_d, of a type 2N247 plane alloy-junction germanium transistor has been determined from measurements of the external short-circuit current gain, by taking account of the effects of the emitter and collector depletion-layer capacitances and the ohmic base resistance. Measurements have been made up to 105 Mc/s and over a range of emitter current from 50 μA to 8 mA. It has been found that the loci of α_d may be interpreted in terms of Kroemer's theoretical treatment, which is based on the existence of a uniform drift field across the base. The value of $\Delta V/kT$ is found to be 5 at room temperature.

LIST OF PRINCIPAL SYMBOLS

α, α_d = External and internal short-circuit current gains, respectively, for common-base connection.
α_0, α_{d0} = Zero-frequency values of α, α_d.
$(\beta_d)_{l.f.}$ = Low-frequency value of the common-emitter internal current gain.
C_c = Collector depletion-layer capacitance.
C_e, C_{e0} = Emitter depletion-layer capacitance and its value for zero emitter bias.
D_n, D_p = Diffusion constants for electrons and holes, respectively.
ΔV = Internal difference of potential energy between emitter and collector.
$f_\alpha, f_{\alpha d}$ = Cut-off frequencies of the external and internal common-base current gains.
G = Emitter input hole conductance of the internal transistor at zero frequency.
I_e = Emitter direct current.
k = Boltzmann's constant.
l_n, l_p = Diffusion lengths of electrons and holes, respectively.
n_e = Equilibrium concentration of electrons in the emitter region.
N = Local donor density in the base.
p = Local hole density in the base.
p_0 = Thermal-equilibrium hole density at the emitter side of the base.
e = Electronic charge.
$r_{bb'}$ = Ohmic base resistance.
$t_0 \equiv \dfrac{4w_0^2}{D_p}\left(\dfrac{kT}{\Delta V}\right)^2$
T = Absolute temperature.
τ_n, τ_p = Lifetimes of electrons in the emitter region and holes in the base, respectively.
V_e, V_c = Emitter-to-base and collector-to-base voltages, respectively.
$\phi_\alpha, \phi_{\alpha d}$ = Phase angles of α and α_d at f_α and $f_{\alpha d}$ respectively.
w_0 = Active base width.
$\omega \equiv 2\pi f$ = Angular frequency.
$x \equiv \dfrac{\omega w_0^2}{2D_p}$
$x_{\alpha d} \equiv \dfrac{\omega_{\alpha d} w_0^2}{2D_p}$, where $\omega_{\alpha d} = 2\pi f_{\alpha d}$.
y_i, y_r, y_f, y_o = Common-base admittance parameters of the internal transistor.
Y_n = Internal emitter input admittance arising from the flow of emitter electron current.

(1) INTRODUCTION

In a series of papers, Kroemer[1–3] has developed a theory of the *p–n–p* 'drift' transistor. For such a transistor the transport of holes across the base is governed primarily by drift in an electric field rather than by diffusion. The theory is based on the assumption of an exponentially decreasing concentration of donor impurities across the base from emitter to collector, so giving rise to a uniform drift field. The construction of drift transistors to date has been based on solid-phase-diffusion techniques.[4–6] In the simplest case, in which only one type of donor atom is diffused into the base, the resulting distribution of donors follows a 'complementary error function' rather than an 'exponential' law. Because an analysis of transistor performance in terms of this practical distribution will be more complicated than if an exponential distribution is assumed, it is expedient to attempt to assess the behaviour of drift transistors in terms of the simplified model. This may be regarded as a first-order approximation to the practical device.

In this connection, the schematic transistor representation of Fig. 1 is useful. The 'internal' transistor is that part of the

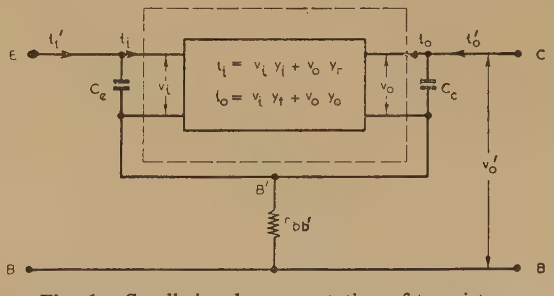


Fig. 1.—Small-signal representation of transistor.
Internal transistor is within dashed lines.

diagram within the dashed lines. The 'external' parameters are C_e, C_c and $r_{bb'}$, the emitter and collector depletion-layer capacitances, and the ohmic base resistance, respectively.

The purpose of the present work is to investigate experimentally the dependence on frequency and emitter current of the internal current gain, α_d ($\equiv [-i_o/i_i]_{v_o=0}$, where i_o, i_i and v_o are defined in Fig. 1), of a type 2N247 *p–n–p* germanium transistor,[6] and to interpret the results on the basis of Kroemer's

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.
Mr. Hyde is now in the Department of Electronic Engineering, University College of North Wales.

theory. To this end, measurements of the 'external' current gain $\alpha\,(\equiv -i_o'/i_i')$, under near-short-circuit conditions, have been made in the frequency range 1–105 Mc/s, using apparatus which has already been described;[7] values of α_d have subsequently been determined by taking account of the effect of C_e, C_c and $r_{bb'}$ on the relationship between α and α_d. In addition, measurements of α_d have been made at a frequency of 50 kc/s, using a bridge method,[8] the results being interpreted in a manner similar to that described by the author for diffusion-type transistors.[9]

(2) THEORETICAL CONSIDERATIONS

(2.1) Admittance Parameters

The admittance parameters are given by Kroemer's theory and may be written as follows:

$$y_i = \frac{G}{2}(1 + X\coth\psi) + Y_n \rightarrow \frac{G}{2}(1 + X) + Y_n \quad . \quad (1a)$$

$$y_r = -\frac{GX}{2K}\operatorname{cosech}\psi \rightarrow -\frac{GX}{K}\varepsilon^{-\psi} \quad . \quad . \quad . \quad . \quad . \quad (1b)$$

$$y_f = -\frac{GX}{2}\varepsilon^{\psi/X}\operatorname{cosech}\psi \rightarrow -GX\varepsilon^{-(\psi/X)(X-1)} \quad . \quad . \quad (1c)$$

$$y_0 = \frac{G}{2K}\varepsilon^{\psi/X}(X\coth\psi - 1) \rightarrow \frac{G}{2K}\varepsilon^{\psi/X}(X - 1) \quad . \quad . \quad (1d)$$

The expressions on the left are exact: those on the right are valid approximations when the internal drift potential ΔV of a particular transistor is greater than $4kT$. The symbols used in eqns. (1a)–(1d) have the following significance:

$$X = \left[1 + \left(\frac{2w_0kT}{l_p\Delta V}\right)^2 + j4\omega\frac{w_0^2}{D_p}\left(\frac{kT}{\Delta V}\right)^2\right]^{1/2} \quad . \quad (1e)$$

$$\frac{\psi}{X} = \frac{\Delta V}{2kT} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1f)$$

$$\frac{1}{K} = \frac{\zeta kT}{ew_0}\frac{\partial w_0}{\partial V_c}\varepsilon^{-eV_e/kT}[(\varepsilon^{eV_e/kT} - 1)\operatorname{cosech}\zeta$$
$$+ \varepsilon^{\Delta V/2kT}(\coth\zeta - \Delta V/2\zeta kT)] \quad . \quad (1g)$$

with $$\zeta = \left[\left(\frac{\Delta V}{2kT}\right)^2 + \left(\frac{w_0}{l_p}\right)^2\right]^{1/2}$$

$$G = \frac{A_e\Delta Ve^2}{(kT)^2}\frac{D_pp_0}{w_0}\varepsilon^{eV_e/kT}$$
$$\simeq eI_e/kT \text{ for } Y_n/G,\ 2w_0kT/l_p\Delta V \ll 1 \ll \Delta V/kT \quad . \quad . \quad (1h)$$

$$\frac{Y_n}{G} = \frac{D_nn_ew_0kT}{D_pl_pp_0\Delta V}(1 + j\omega\tau_n)^{1/2} \quad . \quad . \quad . \quad . \quad (1i)$$

where $\partial w_0/\partial V_c$ is the rate of change of base width with applied collector-to-base voltage, A_e is the emitter area, and $l_p = (D_p\tau_p)^{1/2}$.

(2.2) Internal Short-Circuit Current Gain, α_d

In terms of the expressions given above for y_i and y_f,

$$\alpha_d \equiv -\frac{y_f}{y_i} = \frac{(GX/2)\varepsilon^{\psi/X}\operatorname{cosech}\psi}{(G/2)(X\coth\psi + 1) + Y_n} \rightarrow \frac{GX\varepsilon^{-(\psi/X)(X-1)}}{G(1+X)/2 + Y_n} \quad . \quad . \quad . \quad . \quad (2)$$

If Y_n at zero frequency is very much less than G (high emitter efficiency) and $\omega\tau_n$ is very much less than unity (a usual assumption), Y_n can be ignored in the denominator, and eqn. (2) consequently simplified to

$$\alpha_d = \frac{X\varepsilon^{\psi/X}\operatorname{cosech}\psi}{(X\coth\psi + 1)} \rightarrow \frac{2X}{1+X}\varepsilon^{-(\psi/X)(X-1)} \quad . \quad . \quad (2$$

Theoretical curves in the complex α_d plane, based on eqn. (2 are presented in Fig. 2, curves (a)–(c), for the limiting conditi

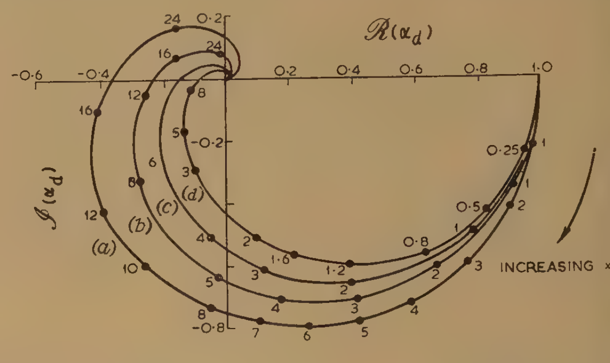


Fig. 2.—Loci of internal current gain of drift and diffusion transisto for $w_0/l_p \rightarrow 0$.
(a) Drift; $\Delta V/kT = 8$.
(b) Drift; $\Delta V/kT = 4$.
(c) Drift; $\Delta V/kT = 2$.
(d) Diffusion; parameter $x \equiv \omega w_0^2/2D_p$

$w_0/l_p = 0$. The curves, which are plotted in terms of t dimensionless parameter $x \equiv \omega w^2/2D_p$, show the dependence α_d on frequency. For curve (a) $\Delta V/kT$ is 8, which is the approx mate upper limit achievable in a germanium transistor. F curves (b) and (c) $\Delta V/kT$ is 4 and 2, respectively. The locus α_d for a pure-diffusion transistor is shown as curve (d) for com parison. This was calculated from the theoretical expression $\alpha_d = \operatorname{sech}\sqrt{(j2x)}$, which applies for $w_0/l_p = 0$.

Curve (a) is an outer limiting locus within which all loci mu lie. The effects of non-unity emitter efficiency (non-zero Y and finite values of w_0/l_p are to make the locus lie slightly insi the limiting curve for the particular value of $\Delta V/kT$ being co sidered. For the low Y_n and w_0/l_p of well-designed transisto these effects are only slight at low frequencies and decrease wi increasing frequency.

A standard parameter, which is used to specify the hig frequency performance of a transistor, is the frequency $f_{\alpha d}$ which the modulus of α_d is reduced to $1/\sqrt{2}$ times α The variation of the corresponding value of x, namely $x_{\alpha d}$ $\pi f_{\alpha d}w_0^2/D_p$, with $\Delta V/kT$ is shown as curve (i) in Fig. 3(It may be seen that curve (ii), representing $x_{\alpha d} = \Delta V/k$ is a useful approximation to the accurate curve, in particul for the values of $\Delta V/kT$ which are realized in practic In this connection it may be noted that the approxima form for $f_{\alpha d}$ suggested by Kroemer,[2,3] which may be p into the form $x_{\alpha d} = (\Delta V/2kT)^{3/2}$, is a poorer approximati than the linear relation $x_{\alpha d} = \Delta V/kT$ suggested above. In a case Kroemer's approximation is not strictly valid because it based on a non-standard definition of $f_{\alpha d}$. For a pure-diffusi transistor, $f_{\alpha d}$ is a sufficient specification of the high-frequen transmission properties of the internal transistor, because t phase angle $\phi_{\alpha d}$ of the current gain at frequency $f_{\alpha d}$ is independe of the dimensions and parameters of the base. Figs. 2 and 3 show that for other than pure-diffusion transistors $\phi_{\alpha d}$ is function of $\Delta V/kT$, so that ideally this parameter should al be specified in addition to $f_{\alpha d}$. A close approximation curve (i) in Fig. 3(b) is given by the dashed curve (ii) representi $\phi_{\alpha d} = 58° + 6\Delta V/kT$.

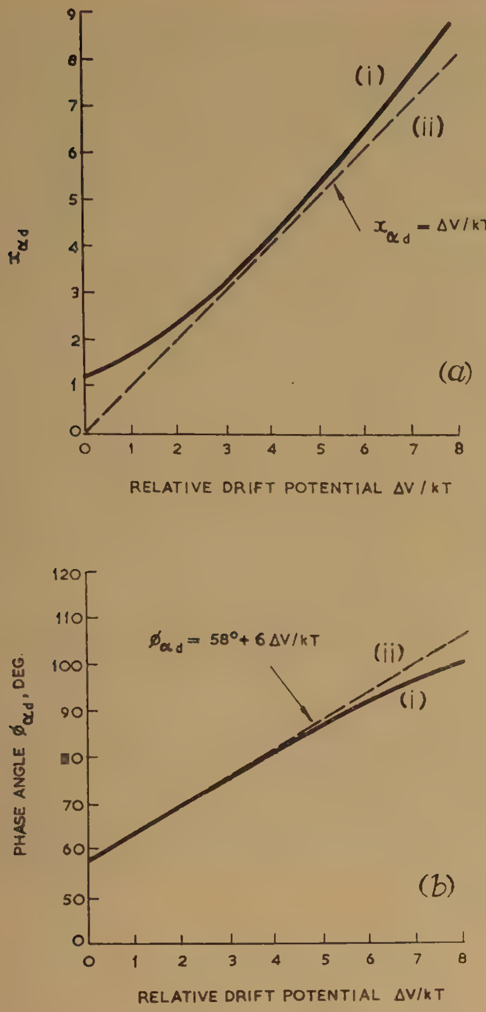


Fig. 3.—Dependence of $x_{\alpha d}$, $\phi_{\alpha d}$ on $\Delta V/kT$.

(a) (i) Theoretical form, $x_{\alpha d} = \omega_{\alpha d} w_0^2/2D_p$.
(ii) Empirical linear approximation $x_{\alpha d} = \Delta V/kT$.
(b) (i) Theoretical form, $\phi_{\alpha d}$.
(ii) Empirical linear approximation, $\phi_{\alpha d} = 58° + 6\Delta V/kT$.

(2.3) Relationship between the External and Internal Short-Circuit Current Gains

If the external current gain is measured by comparing the r.f. voltages developed across two equal small resistances R in the emitter and collector circuits,[7] then in addition to the influence of C_e, C_c and $r_{bb'}$ that of R must also be considered in determining the relationship between α and α_d.

For practical purposes, using Fig. 1, this may be shown to be[7, 9]

$$\alpha = \frac{\dfrac{\alpha_d}{1 + j\omega C_e/y_i} + j\omega C_c r_{bb'}}{1 + j\omega C_c(r_{bb'} + R)} \quad . \quad . \quad . \quad . \quad (3)$$

because terms involving y_r and y_o can be ignored.

As for the diffusion transistor the effect of C_e is expected to be marked at (a) low emitter currents (small G), (b) very high emitter currents (large C_e) and (c) high frequencies ($\omega C_e/y_i$ increases as $\sqrt{\omega}$ at high frequencies). The effect of C_c is allied with that of $r_{bb'}$ and R, and will be more marked the smaller is V_c (large C_c). In general, corrections will have to be made to α in accordance with eqn. (3) to yield α_d.

In this connection C_c and $r_{bb'}$ may be determined relatively easily. The appropriate values of C_e for each I_e may be deduced by extrapolation from measurements made with reverse emitter bias. The basis of this extrapolation is the theoretical expression for C_e,[11] namely $C_e = C_{e0}[V_0/(V_0 - V_e)]^{1/2}$, where C_{e0} is the emitter depletion-layer capacitance for zero emitter bias and V_0 is a constant which depends on the base resistivity. The direct measurement of y_i presents considerable difficulty in the frequency range in question. The inductance of the connecting leads cannot be ignored, while the susceptance of the emitter depletion-layer capacitance is large compared with the susceptive part of y_i for the drift transistor. It is possible, however, to calculate y_i from the theoretical expression of eqn. (1a), provided that Y_n can be ignored and G and X are known. The measurement of G, which is the zero-frequency emitter input conductance, is straightforward.[12] To determine X from

$$X \simeq \left[1 + j4\omega\left(\frac{w_0^2}{D_p}\right)\left(\frac{kT}{\Delta V}\right)^2\right]^{1/2} \equiv (1 + j\omega t_0)^{1/2}$$

it is necessary to determine t_0. Reference to eqn. (5b) in Section 2.4 will show that $t_0 = 4(CR_2 - C_e/G)/[(\Delta V/kT) - 1]$, where C, R_2, C_e and G are known. It remains therefore to find $\Delta V/kT$. This may be done by curve-fitting the loci of α_d to the family of theoretical curves, some of which are illustrated in Fig. 2, at high values of I_e. In these circumstances α_d does not differ appreciably from α, corrected for R, $r_{bb'}$ and C_c, except at the highest frequencies, because $|\omega C_e/y_i|$ is small compared with unity, and corrections involving $(1 + j\omega C_e/y_i)$ are not, therefore, critically dependent on a precise knowledge of y_i.

(2.4) Internal Short-Circuit Current Gain at Low Frequencies

For a diffusion-type transistor it is possible to determine the characteristic parameters of the base experimentally from an analysis of low-frequency measurements of the internal current gain.[9, 13] It is shown below that, for the drift transistor, useful information concerning these parameters may also be obtained in this way. Using the low-frequency expansion of eqn. (2) it may be shown that the low-frequency value (defined by $\omega\tau_p \ll 1$) of the common-emitter current gain $(\beta_d)_{l.f.}$ is given by

$$\left(\frac{1}{\beta_d}\right)_{l.f.} = \frac{1 - \alpha_d}{\alpha_d} = \frac{kT}{\Delta V}\left\{\left[\frac{w_0^2}{l_p^2}\left(1 - \frac{kT}{\Delta V}\right) + \frac{D_n n_e w_0}{D_p l_n p_0}\right] + j\omega\tau_p\frac{w_0^2}{l_p^2}\left(1 - \frac{kT}{\Delta V}\right)\right\} \quad . \quad (4)$$

An analysis of the balance conditions of the low-frequency bridge, developed by Evans[8] for measuring $(1 - \alpha)/\alpha$, yields

$$\frac{R_2}{R_1} = \mathscr{R}\left(\frac{1}{\beta_d}\right)_{l.f.} \simeq \frac{kT}{\Delta V}\left[\frac{w_0^2}{l_p^2}\left(1 - \frac{kT}{\Delta V}\right) + \frac{D_n n_e w_0}{D_p l_n p_0}\right] \quad . \quad (5a)$$

$$CR_2 - \frac{C_e}{G} = \frac{1}{\omega}\mathscr{I}\left(\frac{1}{\beta_d}\right)_{l.f.} \simeq \frac{w_0^2}{D_p}\left(\frac{kT}{\Delta V}\right)\left(1 - \frac{kT}{\Delta V}\right) \quad . \quad (5b)$$

$$CR_1\left(1 - \frac{C_e}{GCR_2}\right) = \frac{\dfrac{1}{\omega}\mathscr{I}\left(\dfrac{1}{\beta_d}\right)_{l.f.}}{\mathscr{R}\left(\dfrac{1}{\beta_d}\right)_{l.f.}} \simeq \frac{\tau_p\left(\dfrac{w_0^2}{l_p^2}\right)\left(1 - \dfrac{kT}{\Delta V}\right)}{\dfrac{w_0^2}{l_p^2}\left(1 - \dfrac{kT}{\Delta V}\right) + \dfrac{D_n n_e w_0}{D_p l_n p_0}} \quad . \quad . \quad . \quad . \quad (5c)$$

where R_1, R_2 and C are parameters of the bridge.[9] The significance of the three relationships of eqn. (5) is as follows:

(a) *Eqn.* (5a).—R_1/R_2 is approximately the base-to-collector current gain, β_{d0}, at zero frequency.

(*b*) *Eqn.* (5*b*).—This may be used to derive an approximate value of $f_{\alpha d}$. It has been shown in Section 2.2 that $x_{\alpha d}$ is given quite closely by $\Delta V/kT$, so that $f_{\alpha d} \simeq (D_p/\pi w_0^2)(\Delta V/kT)$. In terms of $CR_2 - C_e/G$ we may therefore write

$$f_{\alpha d} = \frac{1}{\pi}\frac{1 - kT/\Delta V}{(CR_2 - C_e/G)} \simeq \frac{0{\cdot}8}{\pi(CR_2 - C_e/G)} \quad . \quad . \quad (6)$$

The coefficient 0·8 arises if $\Delta V/kT$ is taken as 5. The error introduced by the use of this fixed coefficient will not be more than about 10% for practical values of $\Delta V/kT$ which are likely to arise.

(*c*) *Eqn.* (5*c*).—It is apparent that $CR_1(1 - C_e/GCR_2)$ is approximately equal to τ_p if the emitter efficiency is high, i.e. if $D_n n_e w_0/D_p l_n p_0 \ll (w_0^2/l_p^2)(1 - kT/\Delta V)$. If this inequality is not valid then it may at least be inferred that τ_p is greater than $CR_1(1 - C_e/GCR_2)$.

(2.5) Departures from the Low-Level One-Dimensional Theory

The general theoretical treatment, on which the whole of the foregoing discussion has been based, was carried out for the following conditions:

(*a*) The hole lifetime, τ_p, is constant throughout the base region.
(*b*) The surface of the base plays no part in controlling the transport of holes between emitter and collector.
(*c*) No recombination or avalanche multiplication occurs within the collector depletion layer.
(*d*) The transit time, τ_c, for holes crossing the collector depletion layer is negligible compared with that through the active base region.
(*e*) The diffusion constant and mobility of holes in the base have their low-level values, i.e. those which apply when the local density p of holes in the base is very much less than the equilibrium concentration N of electrons at each point in the base.

In practice τ_p is unlikely to be constant across the base: the work of Shockley and Read[14] and Burton *et al.*[15] has shown that lifetime is a function of resistivity. Furthermore τ_p is also likely to be dependent on p/N,[14,16] which is itself not constant across the base. The effect of a non-uniform and injection-level-dependent τ_p will be to produce some variation in α_d at low frequencies, but little variation at high frequencies, because then the frequency distribution around the theoretical loci of α_d is determined solely by $w_0^2/2D_p$ [see eqns. (2) and (1*e*)].

Webster[17] showed that the effect of an active recombination surface ring surrounding the emitter was to produce a decrease in α_{d0}, while from Schrieffer's[18] work on surface scattering it is apparent that any holes scattered by the base surface will have their diffusivity decreased. These effects, as that of avalanche multiplication, are likely to be small in alloy-type transistors having plane geometry,[6,19] but in any case will have little influence on α_d at high frequencies.

Kroemer[3] has calculated that τ_c is not significant compared with the transit time across the active base region for transistors whose $f_{\alpha d}$ is less than 100 Mc/s (such as the type 2N247). He has also shown that at high injection levels the mobility of holes along the drift field is effectively reduced to $N/(N + p)$ times that at low level, while at the same time the diffusion constant for the holes is increased to $(N + 2p)/(N + p)$ times that at low level, as for the diffusion transistor.[17,20] The effect of this behaviour on the loci of α_d will be to cause a transition from the low-level 'drift' curve [lying on or inside curve (*a*) of Fig. 2] to the unique 'pure-diffusion' curve [curve (*d*) of Fig. 2], but with a frequency distribution determined in the limit, as p tends to infinity, by $w_0^2/4D_p$ rather than by $w_0^2/2D_p$. A further manifestation of high-level injection is the fall in emitter efficiency which then occurs:[17,20] this is equivalent to an effective increase of the term $D_n n_e w_0/D_p l_n p_0$.

(3) MEASUREMENTS

All measurements were made at room temperature (22 ± 2° and, unless otherwise stated, with a collector-to-base voltage −9 volts. The basic measurement made was that of the nea short-circuit current gain in the frequency range 1–105 Mc with $R = 50$ ohms. The emitter current was varied from 50 μ to 8 mA, and care was taken to ensure that 'small-signal' co ditions prevailed. It was established that no significant erro were introduced by unavoidable stray capacitances; this w done by connecting 0·5 pF capacitors between the emitter a base and between the emitter and collector. No changes the readings were noted up to the highest frequencies as a res of this. This was to be expected,[7] in view of the low value the product $r_{bb'}C_c$ for the transistor.[6]

So that α_d could be determined from the above readings, accordance with eqn. (3), the parameters C_e, $r_{bb'}$ and C_c we measured, while y_i was derived from measurements of G and The product $C_c r_{bb'}$ was determined by h-type neutralization[21] the transistor at 10 Mc/s and was found to be 50×10^{-12} s The ohmic base resistance, $r_{bb'}$, was measured as 26 ohms z-type neutralization[22] at 10 Mc/s, with $V_c = -1$ volt, and was therefore deduced to be 1·9 pF.

In view of the critical effect of C_e on the relationship betwe α and α_d in some circumstances, measurement of this parame was carried out by two methods under conditions of reve emitter bias. The first was a direct admittance-bridge measu ment of the emitter-to-base input capacitance at 4 Mc/s, at wh frequency the contribution of the ohmic base resistance to readings was insignificant: this measurement was made with collector biased normally. The second was a determination the product $C_e r_{bb'}$, as for $C_c r_{bb'}$ above, with the collector un forward bias and acting as an emitter. The linear plot of 1

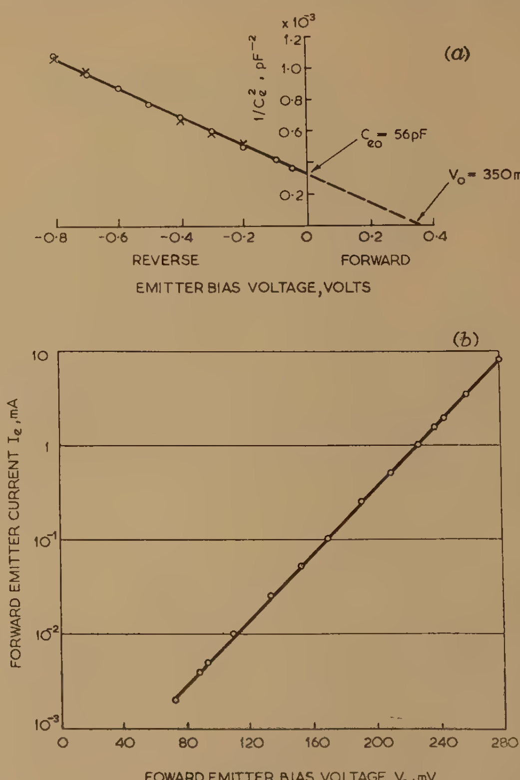


Fig. 4.—Emitter measurements.

(*a*) Dependence of $1/C_e^2$ on emitter voltage V_e.
⊙ Admittance-bridge measurement.
× Data from $C_e r_{bb'}$ measurement.
(*b*) Static V_e/I_e characteristic of emitter-to-base junction.

versus V_e for reverse emitter bias up to $-0{\cdot}8$ volt is shown in Fig. 4(*a*): it may be noted that a single curve fits both sets of experimental data. From this curve the characteristic parameters may be seen to be $C_{eo} = 56$pF and $V_0 = 0{\cdot}35$ volt. The forward d.c. characteristic of the emitter, relating I_e and V_e, is shown in Fig. 4(*b*). The slope of this characteristic corresponds to a value of $39{\cdot}5$ volt^{-1} for kT/e. Using Fig. 4(*b*) in conjunction with Fig. 4(*a*), C_e may be calculated for any value of I_e.

The low-frequency base-to-collector current gain was measured[8] at 50kc/s over the range of emitter current from $10\,\mu$A to 8mA. G, which is involved in conjunction with C_e in the analysis of these low-frequency data and also in the determination of X, was measured by the method described by Boothroyd and Almond.[12] It was found to be given by $G = eI_e/kT$, as was expected theoretically.

The following auxiliary measurements were made:

(*a*) *The thermal resistance of the collector region.*—The electronic method, described by Loofbourrow and Ollendorf[23] for the measurement of the collector saturation current, I_{co}, was used and a figure of $0{\cdot}25°$C/mW deduced.

(*b*) *The reverse d.c. characteristic of the emitter junction.*—The saturation current was found to be less than $0{\cdot}2\,\mu$A and the junction leakage resistance greater than 2 megohms.

(4) RESULTS

In Figs. 5(*a*) and (*b*), curves (i) show the dependence on I_e of the 3dB-down frequency f_α and corresponding phase angle ϕ_α

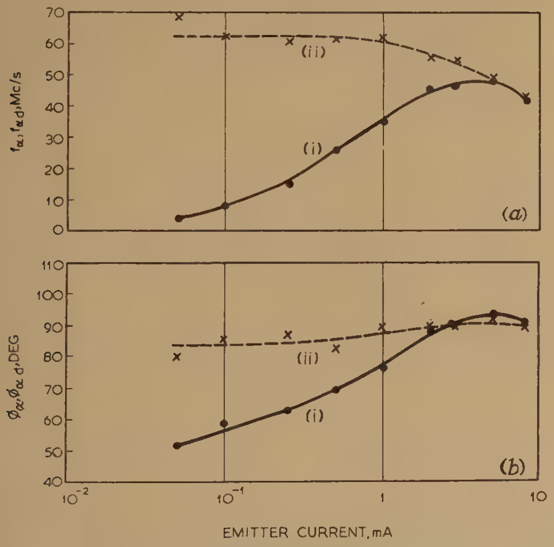


Fig. 5.—Dependence on emitter current of f_α, $f_{\alpha d}$, ϕ_d and $\phi_{\alpha d}$.
(*a*) —— f_α; ----- $f_{\alpha d}$
(*b*) —— ϕ_α; — — — $\phi_{\alpha d}$

of the near-short-circuit external current gain. The individual experimental values (marked on the diagram) of these parameters were derived from complete loci of α covering the frequency range 50kc/s–105Mc/s. Four of these, for emitter currents of $100\,\mu$A, $500\,\mu$A, 2mA and 8mA, are shown as the full curves in parts (i) of Figs. 6(*a*)–(*d*) respectively. In the same diagrams $\alpha_d/(1 + j\omega C_e/y_i)$, derived from α by taking account of R, $r_{bb'}$ and C_c at each frequency, and α_d derived from these by multiplying them by $(1 + j\omega C_e/y_i)$, are shown. The dashed curves represent the loci of α_d. From these loci $f_{\alpha d}$ and $\phi_{\alpha d}$ have been determined and are plotted as crosses in Figs. 5(*a*) and (*b*), smooth dashed curves being drawn through them. In parts (ii) of Fig. 6 the loci of α_d are repeated and compared with theoretical curves of the family illustrated by Fig. 2. The data obtained from the low-frequency bridge measurements of β_d are plotted in Fig. 7, as functions of I_e.

(5) INTERPRETATION OF RESULTS

(5.1) Low-Frequency Bridge Data

(*a*) *Fig.* 7(*a*), *eqn.* (5*a*).—It is evident that the zero-frequency value of β_d increases monotonically with I_e. Since no peak has been reached and the overall variation is some ten to one, it is clear that the term $(D_n n_e w_0/D_p l_n p_0)'$,* the expected increase of which with increasing I_e would cause a decrease of the current gain, is insignificant compared with $(w_0^2/l_p'^2)(1 - kT/\Delta V)$, at least for small I_e. The rise may therefore be ascribed to an increase of the effective diffusion length of holes in the base with increasing injection level.

(*b*) *Fig.* 7(*b*), *eqn.* (5*b*).—At low I_e the term C_e/G accounts almost entirely for the measured value of CR_2. Because of this, no real significance should be attached to the absolute values of $CR_2 - C_e/G$ in this region. The fact that the smooth, dashed curve representing this quantity is almost invariant up to the highest emitter currents, which is the expected behaviour of the right-hand side of eqn. (5*b*) in the absence of high-injection-level effects and a marked temperature rise of the base (that of the collector region is only 18°C at $I_e = 8$mA), is strongly suggestive that the correction term C_e/G is in fact appropriate. For I_e greater than a few hundred microamperes the contribution of C_e/G to the left-hand side of the equation becomes small. For $I_e = 2$mA, $CR_2 - C_e/G$ is $0{\cdot}0043$ microsec. The use of this value in eqn. (6) yields $f_{\alpha d} = 59$ Mc/s. For the highest emitter currents a slight rise of $(CR_2 - C_e/G)$ was observed—for I_e equal to 8mA it was $0{\cdot}0051$ microsec. This rise could result from a combination of the following effects: (*a*) a decrease of D_p with the increase of temperature which accompanies the higher power dissipation involved, (*b*) the corresponding increase of $kT/\Delta V$ and (*c*) the onset of high-level-injection effects at the collector side of the base (see Section 2.5). Because the detailed variation of N between emitter and collector is unknown, it is not possible to ascribe a unique temperature dependence to D_p or to take account of injection level on a rigorous basis. It is not possible, therefore, to interpret the rise of $(w_0^2/D_p)(kT/\Delta V)(1 - kT/\Delta V)$ analytically. Because the overall change in this quantity is small between low and high values of I_e, it is clear that the increase of hole diffusion length in the base referred to in (*a*) must be due predominantly to an increased effective hole lifetime.

(*c*) *Fig.* 7(*c*), *eqn.* (5*c*).—In (*a*) it was pointed out that $(D_n n_e w_0/D_p l_n p_0)'$ could be ignored compared with $(w_0^2/l_p'^2)$ $(1 - kT/\Delta V)$ at small emitter currents. In these circumstances $CR_1(1 - C_e/GCR_2)$ may be interpreted as the lifetime of holes in the base. This is seen to rise monotonically with increasing I_e and is approximately a replica of the variation of R_1/R_2 because of the almost constant value of $CR_2 - C_e/G$. It is an open question whether the rise in $CR_1/(1 - C_e/GCR_2)$ represents a real increase of bulk lifetime or only an effective increase of this parameter which results from a proportionate decrease of holes recombining on the surface of the base as I_e is increased.

(5.2) Current-Gain Loci

The procedure for correcting α for R, $r_{bb'}$, and C_c at each frequency is straightforward.[7] The further step of correcting for $\omega C_e/y_i$ is complicated by the necessity of knowing the value of $\Delta V/kT$ to be used in the calculation.

* Primes are used to denote 'effective' values, a dependence on injection level being implied.

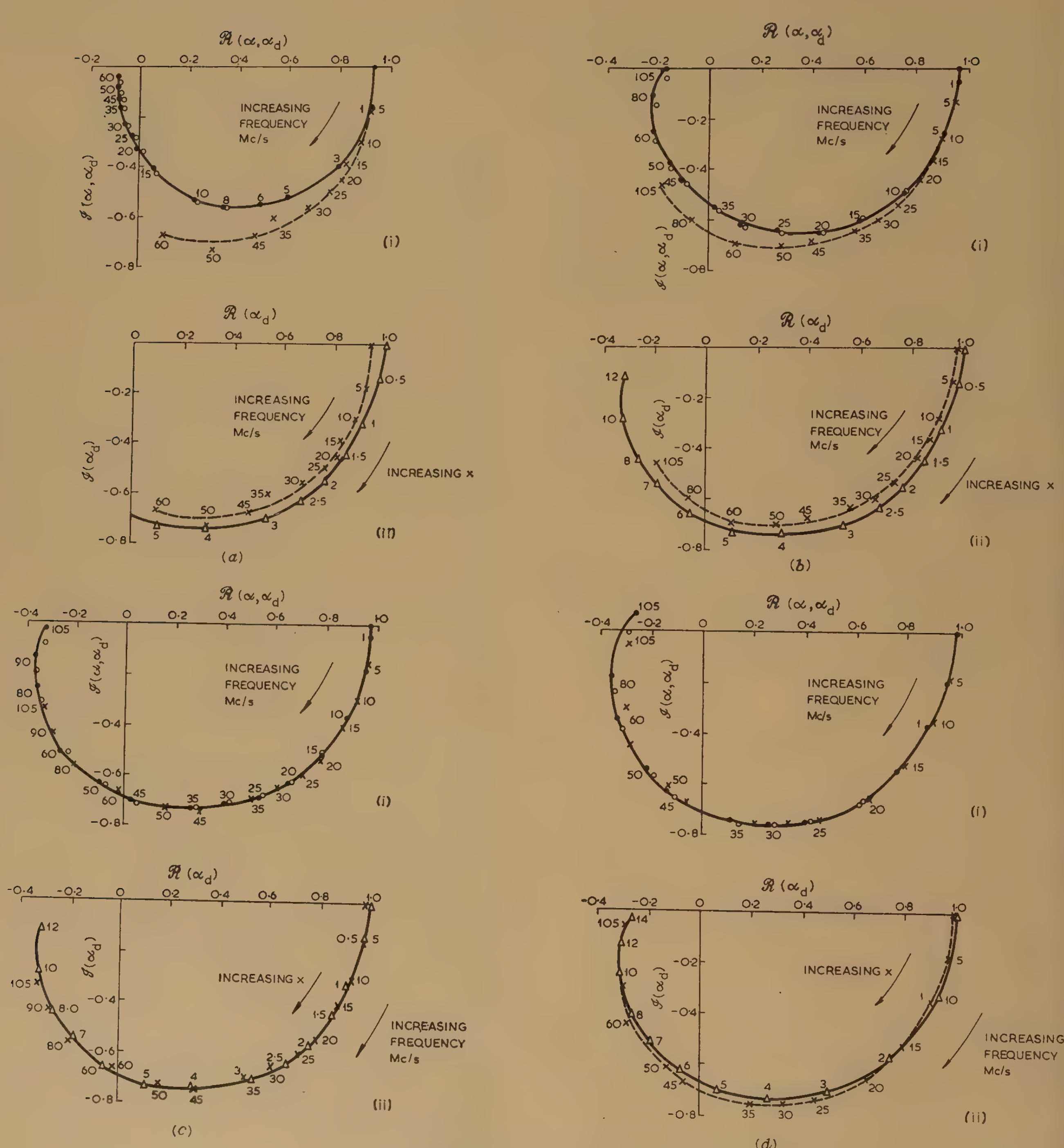


Fig. 6.—Dependence of current gain on frequency.

(i) —●—●—●— External short-circuit current gain α. ○ α corrected for R, r_{bb}' and C_e. —×——×— α_d

(ii) Comparison of experimental (—×——×—) and theoretical (—△—△—) limiting forms of α_d for $\Delta V/kT = 5$ in (a)–(c) and for $\Delta V/kT = 4{\cdot}75$ in (d).
Parameters: $V_c = -9$ volts, $T = 22°$ C.

(a) $I_e = 100\,\mu$A. (b) $I_e = 500\,\mu$A. (c) $I_e = 2$ mA. (d) $I_e = 8$ mA.

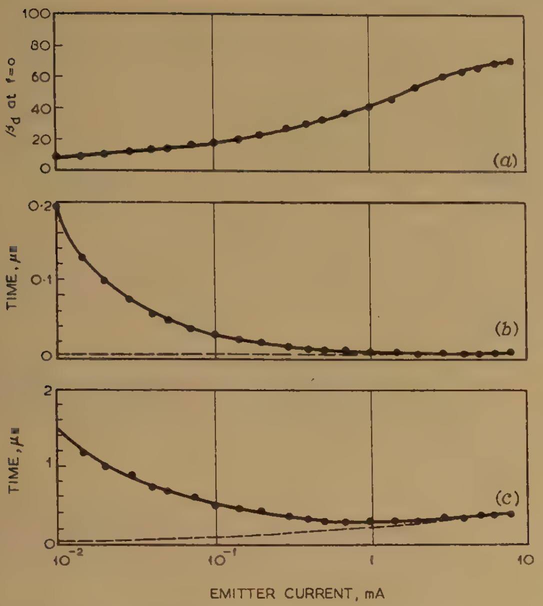


Fig. 7.—Low-frequency bridge data.

(a) Variation of $\beta_d|_{f=0} = 0$ with I_e.
(b) —— Variation of CR_2 with I_e.
---- Variation of $CR_2 - C_e/G = (1/\omega)\mathscr{I}(1/\beta_d)_{l.f.}$ with I_e.
(c) —— Variation of CR_1 with I_e.
---- Variation of $1 - C_e/GCR = (1/\omega)\mathscr{I}(1/\beta_d)_{l.f.}/\mathscr{R}(1/\beta_d)_{l.f.}$ with I_e.

$\Delta V/kT$ was estimated by finding the particular curve of α_d, of the family illustrated by Fig. 2, which agreed closely with the experimental short-circuit current-gain locus at $I_e = 2$ mA. At this current the temperature rise of the base was not significant, and in addition the correction for $\omega C_e/y_i$ was not expected to be large except at the highest frequencies. $\Delta V/kT$ was taken as 5. Then, using 0·0043 microsec for $(CR_2 - C_e/G)$, y_i was calculated as a function of frequency in the manner described in Section 2.3. The resulting locus of α_d for $I_e = 2$ mA was then compared with the theoretical locus for $\Delta V/kT = 5$. This comparison suggested that the value of 5 for $\Delta V/kT$ was in fact appropriate. This was then used in the correction of all experimental data for I_e up to and including 2 mA, while to take account of the temperature rise in the base a round figure of 4·75 was taken for $\Delta V/kT$ for the emitter currents of 5 and 8 mA.

The comparisons of the experimentally derived loci of α_d with the appropriate theoretical loci are shown in Figs. 6(ii). In this connection it should be pointed out that the theoretical loci are calculated on the basis of $w_0/l_p = 0$, so that they meet the real axis at $\alpha_d = 1·0$, whereas the experimental curves meet this axis at α_{d0}. It would be more appropriate to compare the experimental curve with theoretical curves based on a non-zero value of w_0/l_p. This has not been done because (a) there is some uncertainty concerning the interpretation of the hole lifetime at low emitter currents, both as regards its absolute magnitude and also its physical origin as a bulk or part-bulk and part-surface-controlled parameter; and (b) the high-frequency expansion of eqn. (2a) becomes increasingly independent of the real or effective value taken for w_0/l_p as the frequency is increased. The disparity between the experimental and theoretical curves at frequencies less than $f_{\alpha d}$ should be greater the larger w_0/l_p, or, equivalently, the smaller α_{d0}. This was indeed found to be the case. At the smaller emitter currents the experimental curves lie slightly inside the limiting theoretical curves, as is expected. At larger I_e, i.e. 2 and 8 mA, for which α_{d0} approaches unity, there is closer agreement between the curves. In Table 1 the

Table 1

RATIO f/x FOR $I_e = 2$ MA

Frequency f Mc/s	Ratio f/x Mc/s
10	10·5
15	10·9
20	10·5
25	10·9
30	11·1
35	11·2
45	11·2
50	10·6
60	10·3
80	11·4
90	11·2
105	11·3

ratio of f, the frequency defining a point on the experimental curve, to that of the dimensionless parameter $x \equiv \omega w_0^2/2D_p$, which defines the corresponding point on the theoretical curve, is given for $I_e = 2$ mA. The ratio f/x is sensibly constant throughout the whole r.f. range, so that there is agreement not only in shape but also in the frequency distributions of the experimental and theoretical curves. The corresponding f/x calculated from the l.f. bridge data [eqn. (5b)] is 11·8 Mc/s, which agrees well with those of the r.f. range.

In passing from 2 to 8 mA for I_e, there is a shift of the experimental values of α_d to the left, a result of which is a fall in $f_{\alpha d}$ from 56 to 43 Mc/s. There is a concomitant change in the shape of the locus, but this is only slight. The shift may therefore be interpreted in terms of an increase in $w_0^2/2D_p'$ or, equivalently, as a decrease in D_p'. In this connection it is interesting to compare the above values of $f_{\alpha d}$ with those calculated from the first part of eqn. (6), in which $(CR_2 - C_e/G)$ is taken as 0·0043 and 0·0051 microsec and $\Delta V/kT$ as 5 and 4·75, namely 59 and 49 Mc/s, in reasonable agreement with those stated above.

Taking the ratio $f/x = D_p/\pi w_0^2$ as 11 Mc/s at $I_e = 2$ mA and D_p as 47 cm^2/sec leads to $w_0 = 1·2 \times 10^{-3}$ cm. From eqn. (5a) $w_0 = 1·1 \times 10^{-3}$ cm is obtained. These are in reasonable agreement with the mean value of $1·5 \times 10^{-3}$ cm which has seen quoted for the type 2N247 transistor.

The corrections to α for R, $r_{bb'}$, and C_c are small for this transistor at all frequencies, because of the low values of these parameters. The corrections for $\omega C_e/y_i$ are, however, quite large, in particular for low I_e. It may be seen from Fig. 6 that the corrections for $\omega C_e/y_i$ are least for $I_e = 8$ mA. It is possible that under pulsed conditions at much higher currents the effect of C_e will increase again; reference to Fig. 4(a) will show that this will not occur until the bias voltage of the emitter junction approaches 350 mV. That the large corrections involving C_e at small emitter currents give rise to loci of α_d which are appropriately located, slightly within the limiting theoretical loci, and furthermore that the derived values of $f_{\alpha d}$ and $\phi_{\alpha d}$ shown in Fig. 5 are sensibly independent of emitter current, as is expected, is good evidence in support of the validity of the corrections which have been made.

(6) CONCLUSIONS

It would be surprising if perfect agreement had been obtained between theory and experiment in this work because, as was stated in the introduction, the theory is based on an exponential distribution of donors across the base, whereas such a distribution is not expected in practice. The agreement is good enough, however, to suggest that for the particular transistor considered the use of Kroemer's theory is justified. On this basis there is reasonable evidence that a value of 5 for $\Delta V/kT$ is pertinent at room temperature. A very useful result is the good agreement shown between the low-level values of approximately 60 Mc/s

[see Fig. 5(*a*)] and of 59 Mc/s for $f_{\alpha d}$ derived respectively (*a*) from direct measurements of α corrected for R, $r_{bb'}$, C_c and $\omega C_e/y_i$, and (*b*) by an extrapolation from low-frequency data, which is based on the theoretical expression for α_d. If the more general validity of this extrapolation can be established, it will provide a useful means of determining $f_{\alpha d}$ from low-frequency measurements using quite simple apparatus. For the drift transistor, however, a knowledge of $f_{\alpha d}$ is not sufficient to specify the high-frequency transmission properties of the base. It is also necessary to specify the phase angle $\phi_{\alpha d}$ of the current gain at $f_{\alpha d}$ or, equivalently the value of $\Delta V/kT$.

In contrast with the diffusion transistor with its homogeneous base and consequently large $r_{bb'}$, and C_c, these parameters are small for the drift transistor and contribute little to the difference between α and α_d. C_e is comparatively large for the drift transistor, however, and its gross effect on the relationship between α and α_d at low and even moderate emitter currents has been demonstrated.

(7) ACKNOWLEDGMENT

The work described was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(8) REFERENCES

(1) Kroemer, H.: 'The Drift Transistor', *Naturwissenschaften*, 1953, **40**, p. 578.
(2) Kroemer, H.: 'On the Theory of the Diffusion and the Drift Transistor', *Archiv der electrischen Übertragung*, 1954, **8**, Part I, p. 223, Part II, p. 363, and Part III, p. 499.
(3) Kroemer, H.: 'The Drift Transistor', from 'Transistors I' (R.C.A. Laboratories, 1956), p. 202.
(4) Tanenbaum, M., and Thomas, D. E.: 'Diffused Emitter and Base Silicon Transistors', *Bell System Technical Journal*, 1956, **35**, p. 1.
(5) Lee, C. A.: 'A High-Frequency Diffused-Base Germanium Transistor', *ibid.*, 1956, **35**, p. 23.
(6) Kestenbaum, A. L., and Ditrick, N. H.: 'Design, Construction and High-Frequency Performance of Drift Transistors', *R.C.A. Review*, 1957, **18**, p. 12.
(7) Hyde, F. J., and Smith, R. W.: 'An Investigation of the Current Gain of Transistors at Frequencies up to 105 Mc/s', *Proceedings I.E.E.*, Paper No. 2548 R, May, 1958 (**105** B, p. 221).
(8) Evans, D. M.: 'Measurements on Alloy-Type Germanium Transistors and their Relation to Theory', *Journal of Electronics*, 1956, **1**, p. 461.
(9) Hyde, F. J.: 'Some Measurements on Commercial Transi tors and their Relation to Theory', *Proceedings I.E.E.*, Paper No. 2438 R, January, 1958 (**105** B, p. 45).
(10) Pritchard, R. L.: 'Frequency-Variation of Junctio Transistor Parameters', *Proceedings of the Institute Radio Engineers*, 1954, **42**, p. 786.
(11) Harten, H. V., Koch, W., Rath, W. L., and Schultz, W 'Barrier-Layer Capacitance of the Ge–In Alloyed Rect fier', *Zeitschrift für Physik*, 1954, **138**, p. 336.
(12) Boothroyd, A. R., and Almond, J.: 'A Bridge f Measuring the A.C. Parameters of Junction Transistors *Proceedings I.E.E.*, Paper No. 1681 M, September, 19 (**101**, Part III, p. 314).
(13) Hyde, F. J.: 'An Investigation of the Dependence of th Current Gain of a Plane-Alloy-Junction Transistor c Emitter Current and Frequency' (see page 391).
(14) Shockley, W., and Read, W. T., Jnr.: 'Statistics of th Recombination of Holes and Electrons', *Physical Revie* 1952, **87**, p. 835.
(15) Burton, J. A., Hull, G. W., Morin, F. J., and Severien J. C.: 'Effect of Nickel and Copper Impurities on th Recombination of Holes and Electrons in Germanium *Journal of Physical Chemistry*, 1953, **57**, p. 853.
(16) Armstrong, L. D., Carlson, C. L., and Bentivegna, M 'P–N–P Transistors using High-Emitter-Efficiency All Materials', *R.C.A. Review*, 1956, **17**, p. 37.
(17) Webster, W. M.: 'On the Variation of Junction Transist Current-Amplification Factor with Emitter Curren *Proceedings of the Institute of Radio Engineers*, 195 **42**, p. 914.
(18) Schrieffer, J. R.: 'Effective Carrier Mobilities in Surfa Space Charge Layers', *Physical Review*, 1955, **97**, p. 64
(19) Mueller, C. W., and Ditrick, N. H.: 'Uniform Plan Alloy Junctions for Germanium Transistors', fro 'Transistors I' (R.C.A. Laboratories, 1956), p. 121.
(20) Rittner, E.: 'Extension of the Theory of the Junctio Transistor', *Physical Review*, 1954, **94**, p. 1161.
(21) Turner, R. J.: 'Surface-Barrier Transistor Measuremen and Applications', *Tele-Tech*, 1954, **13**, p. 78.
(22) Molozzi, A. R., Page, D. F., and Boothroyd, A. R 'Measurement of High-Frequency Equivalent Circu Parameters of Junction and Surface-Barrier Transistor *Transactions of the Institute of Radio Engineers*, 195 **ED-4**, p. 120.
(23) Loofbourrow, K. E., and Ollendorf, J.: 'Equipment f Measuring Junction Temperature of an Operating Tra sistor', from 'Transistors I' (R.C.A. Laboratories, 195 p. 353.

621.382:621.372.5

The Institution of Electrical Engineers
Paper No. 2857 E
July 1959

HIGH-FREQUENCY POWER GAIN OF THE DRIFT TRANSISTOR

By F. J. HYDE, M.Sc., Associate Member.

(*The paper was first received* 28*th March, and in revised form* 22*nd December*, 1958.)

SUMMARY

Approximate expressions are derived for (*a*) the critical frequency above which the ideal transistor is unconditionally stable, and (*b*) the resulting maximum available gain, for the common-emitter configuration. These involve the internal cut-off frequency, the low-frequency emitter input conductance, the ohmic base resistance and the emitter and collector depletion-layer capacitances.

LIST OF PRINCIPAL SYMBOLS

A = Maximum available gain.
α_d = Internal short-circuit current gain for common-base connection.
$f_{\alpha d}$ = Cut-off frequency of the internal common-base current gain.
C_c = Collector depletion-layer capacitance.
C_e = Emitter depletion-layer capacitance.
D_p = Diffusion constant for holes.
ΔV = Internal difference of potential energy between emitter and collector.
G = Emitter input hole conductance of the internal transistor at zero frequency.
$h_{ie}, h_{re}, h_{fe}, h_{oe}$ = Common emitter hybrid parameters for the complete transistor (including C_e, C_c and $r_{bb'}$).
I_e = Emitter direct current.
k = Boltzmann's constant.
e = Electronic charge.
$r_{bb'}$ = Ohmic base resistance.
T = Absolute temperature.
w_0 = Active base width.
$\omega = 2\pi f$ = Angular frequency.
ω_{crit} = Critical frequency above which the transistor is unconditionally stable.
$\omega_{\alpha d}$ = Internal angular cut-off frequency.
y_i, y_r, y_f, y_o = Common-base admittance parameters of the internal transistor.
$y_\Sigma = y_i + y_r + y_f + y_o = \mathscr{R}(y_\Sigma) + j\mathscr{I}(y_\Sigma)$.

(1) INTRODUCTION

In the design of transistor amplifiers it is an advantage if neutralization (unilateralization)[1] is not required to achieve stability. The purpose of the present paper is to derive expressions for a critical frequency above which the drift (inhomogeneous base) transistor is unconditionally stable in the absence of external feedback and in consequence requires no neutralization, and also expressions for the resulting maximum available gain. The latter will be achieved in practice with conjugate-matched terminations at the input and output. The analysis is approximate but should serve as a guide to the expected behaviour. For common-emitter connection the complete transistor may be represented by the overall h_e parameters* in accordance with eqn. (1), in which the currents and voltages have their usual significance, as illustrated in Fig. 1:

$$\left.\begin{aligned} v_{ie} &= h_{ie}i_{ie} + h_{re}v_{oe} \\ i_{oe} &= h_{fe}i_{ie} + h_{oe}v_{oe} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (1)$$

The critical angular frequency ω_{crit} may be determined from the equality condition of Linvill's criterion for stability,[2,3] namely

$$|h_{re}h_{fe}| + \mathscr{R}(h_{re}h_{fe}) \leqslant 2\mathscr{R}(h_{ie})\mathscr{R}(h_{oe}) \quad . \quad . \quad . \quad (2)$$

in which the h_e parameters are given their frequency-dependent values.

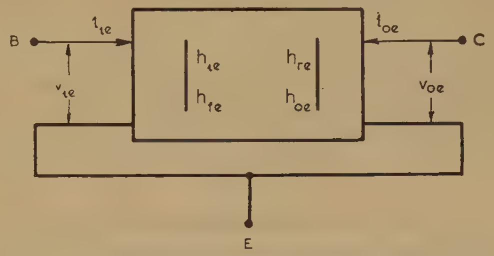


Fig. 1.—Common-emitter representation of the complete transistor.

The maximum available gain A, for frequencies somewhat in excess of ω_{crit}, is given by[2]

$$A = \frac{|h_{fe}|^2}{4\mathscr{R}(h_{ie})\mathscr{R}(h_{oe}) - 2\mathscr{R}(h_{fe}h_{re})} \quad . \quad . \quad . \quad (3)$$

Pritchard[4,5]* has calculated A and ω_{crit} for diffusion-type transistors (transistors having homogeneous-base resistivity) with special reference to the common-emitter configuration which is widely used in practice. Here we are concerned with a *p-n-p* drift transistor having an ideal exponential distribution of donor impurities across the base,[6] and in consequence a uniform drift field. Following Pritchard's calculations the common-emitter configuration will be studied. The parameters arising in eqns. (2) and (3) may be shown, from a consideration of the complete h_e parameters derived in the Appendix, to be given approximately by the following:

$$\mathscr{R}(h_{ie}) = r_{bb'} + \frac{\mathscr{R}(y_\Sigma/G)}{|y_\Sigma|^2} \simeq r_{bb'} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4a)$$

$$\mathscr{R}(h_{oe}) \simeq \frac{\omega C_c}{\mathscr{I}(y_\Sigma/G)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4b)$$

$$|h_{fe}|^2 \simeq \frac{(1 - 2\gamma^2)^2 + (2\gamma + \omega C_c/G)^2}{[\mathscr{I}(y_\Sigma/G)]^2} \quad . \quad . \quad . \quad (4c)$$

$$\mathscr{R}(h_{fe}h_{re}) \simeq \frac{-(\omega C_c/G)(2\gamma + \omega C_c/G)}{[\mathscr{I}(y_\Sigma/G)]^2} \quad . \quad . \quad . \quad . \quad (4d)$$

* The h_e parameters are used here because the ohmic base resistance, $r_{bb'}$, occurs only in one of them, namely h_{ie}, with consequent algebraic simplification.

* Pritchard calculated the maximum available gain for a transistor which was resistance matched at the input and conjugate matched at the output. The approximations made by him, however, are such that his result is equivalent to that derived from eqn. (3) for alloy-type transistors.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.
Mr. Hyde is now in the Department of Electronic Engineering, University College of North Wales.

$$|h_{re}h_{fe}| \simeq \frac{(\omega C_c/G)[(1-2\gamma^2)^2+(2\gamma+\omega C_c/G)^2]^{1/2}}{[\mathscr{I}(y_\Sigma/G)^2]} \quad (4e)$$

where $\gamma = \omega/\omega_{\alpha d}$.

(2) CRITICAL ANGULAR FREQUENCY, ω_{crit}

Substituting into eqn. (2) from eqn. (4) and using the first-order approximation for $\mathscr{I}(y_\Sigma/G)$, namely

$$\mathscr{I}(y_\Sigma/G) \simeq 2\gamma + \omega(C_e + C_c)/G \quad . \quad . \quad . \quad (5)$$

where C_e is the emitter depletion-layer capacitance, it follows that γ_{crit} ($\equiv \omega_{crit}/\omega_{\alpha d}$) is given by

$$\gamma_{crit} = 1/2\Big(\{r_{bb'}G[2+\omega_{\alpha d}(C_e+C_c)/G]+1\}^2 + 2r_{bb'}\omega_{\alpha d}C_c[1+\omega_{\alpha d}(C_e+C_c)/2G]\Big)^{1/2} \quad (6)$$

Because $G \simeq eI_e/kT$, it is clear that, if I_e is considered as the independent variable and $\omega_{\alpha d}$ and C_e are regarded as sensibly constant, ω_{crit} will have a maximum value, which is less than one-half, when the denominator of eqn. (6) is a minimum. This condition is given by

$$4\{r_{bb'}G[2+\omega_{\alpha d}(C_e+C_c)/G]+1\} - \omega_{\alpha d}^2(C_e+C_c)C_c/G^2 = 0 \quad . \quad . \quad . \quad . \quad (7)$$

At large I_e, such that the term $4r_{bb'}G$ in the denominator of eqn. (6) is dominant,

$$\gamma_{crit} \simeq \frac{1}{4r_{bb'}G} \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

(3) THE MAXIMUM AVAILABLE GAIN, A

Substituting from eqns. (4) and (5) into eqn. (3) leads to

$$A = \frac{G^2\omega_{\alpha d} + \omega^2 C_c(4G + \omega_{\alpha d}C_c)}{2\omega^2 C_c\{2G(1+2r_{bb'}G) + \omega_{\alpha d}[2r_{bb'}G(C_e+C_c)+C_c]\}} \quad (9)$$

For large emitter currents this reduces to

$$A \simeq \frac{\omega_{\alpha d}}{8\omega^2 r_{bb'}C_c} \quad . \quad . \quad . \quad . \quad . \quad (10)$$

(4) DISCUSSION

At the larger emitter currents it is clear from eqn. (8) that the larger is $r_{bb'}$ the greater will be the relative frequency range below $\omega_{\alpha d}$ within which the transistor is unconditionally stable for a given value of I_e (note here that G is approximately equal to $I_e/25$ mhos, where I_e is measured in milliamperes). The corresponding expression for A [eqn. (10)] suggests that $r_{bb'}C_c$ should be small. It is important, therefore, to make this product small through C_c rather than $r_{bb'}$, so that a wide range of stable operation is achieved. It is suggested by eqn. (8) that a wider frequency range will be obtained the larger is I_e. Within the limits of low-level theory this is true, but high-injection-level effects result in a decrease of gain through (*a*) a decrease of $\omega_{\alpha d}$, and (*b*) the augmenting of C_c by the internal capacitance of the transistor.

It is of interest to compare the corresponding expressions for the diffusion transistor[4,5] at larger I_e. These are:

$$\gamma_{crit} \simeq \frac{1}{2 \cdot 5 r_{bb'}G} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$$A \simeq \frac{0 \cdot 22\omega_{\alpha d}}{\omega^2 r_{bb'}C_c} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

and are very similar to eqns. (8) and (10) respectively.

At low emitter currents the by-passing effects of the depletio layer capacitances with respect to the internal admittance para meters (whose magnitudes are proportional to I_e) becom important. The full expressions for γ_{crit} and A, eqns. (6) and (9 must then be used. The precise dependence of γ_{crit} and A on will depend on the relative values of the r.f. parameters of th transistor, but in general it may be seen that A will fall as is reduced, while γ_{crit} will first rise and then fall, passing throug a maximum when the condition of eqn. (7) is satisfied. To fi ideas it is useful to compute γ_{crit} and A for a type 2N24 transistor[7] for which it has been established that $\Delta V/kT >$ and the following approximate values of the r.f. parameter have been measured:[8] $C_c \simeq 2$ pF; $C_e \simeq 80$ pF (assume invariant, although it increases slowly with I_e); $r_{bb'} \simeq 25$ ohms and $f_{\alpha d} = \omega_{\alpha d}/2\pi \simeq 60$ Mc/s (assumed invariant, although i decreases slightly for I_e above 2 mA, owing to high-injection level effects). γ_{crit} and A, at $f = 30$ Mc/s, are tabulated belo for emitter currents of 100 μA, 1 mA and 10 mA. The maximu value of γ_{crit} arises in the vicinity of $I_e = 50$ μA and is approx mately one-quarter.

Table 1

DEPENDENCE OF γ_{crit} AND A ON I_e (COMPUTED VALUES)

Emitter current I_e	γ_{crit}	A
mA		dB
10^{-1}	0·25	5
1	0·125	12
10	0·022	14*

* This value will be reduced by the fall in $\omega_{\alpha d}$ which occurs at this current.
Type 2N247 transistor. $f = 30$ Mc/s.

(5) CONCLUSIONS

By making use of the approximation $\omega_{\alpha d} \simeq (\Delta V/kT)(2D_p/w$ for the drift transistor and other simplifications, it has bee possible to derive analytical expressions for the critical frequenc above which the transistor is unconditionally stable, and th resulting maximum available gain when the transistor is cor jugate-matched to both the source and load. The conclusion which may be drawn from the analysis, with regard to commo emitter operation, are (*a*) that in design it is more important t make C_c small than $r_{bb'}$ when aiming for a particular value c $C_c r_{bb'}$, and (*b*) that in operation (exclusive of noise consideration it is better to choose a large emitter current rather than a sma one, provided that there is no deterioration in performance c the internal transistor from high-injection-level effects at th current chosen.

(6) ACKNOWLEDGMENTS

The work described was carried out as part of the programm of the Radio Research Board. The paper is published by pe mission of the Director of Radio Research of the Department c Scientific and Industrial Research.

(7) REFERENCES

(1) STERN, A. P., ALDRIDGE, C. A., and CHOW, W. F.: 'Intern Feedback and Neutralization of Transistor Amplifier *Proceedings of the Institute of Radio Engineers*, 1955, 4 p. 838.

(2) LINVILL, J. G., and SCHIMPF, L. G.: 'The Design of Tetro Transistor Amplifiers', *Bell System Technical Journ* 1956, **35**, p. 813.

(3) STERN, A. P.: 'Considerations on the Stability of Acti Elements and Applications to Transistors', *Institute Radio Engineers Convention Record*, 1956, **4**, Part II, p. 4

(4) PRITCHARD, R. L.: 'High-Frequency Power Gain of Junction Transistors', *Proceedings of the Institute of Radio Engineers*, 1955, **43**, p. 1075.
(5) PRITCHARD, R. L.: 'Measurement Considerations in High-Frequency Power Gain of Junction Transistors', *ibid.*, 1956, **44**, p. 1050.
(6) KROEMER, H.: 'The Drift Transistor', from 'Transistors I' (R.C.A. Laboratories, 1956), p. 202.
(7) KESTENBAUM, A. L., and DITRICK, N. H.: 'Design, Construction and High-Frequency Performance of Drift Transistors', *R.C.A. Review*, 1957, **18**, p. 12.
(8) HYDE, F. J.: 'An Investigation of the Current Gain of a Drift Transistor at Frequencies up to 105 Mc/s' (see page 397).

(8) APPENDICES

(8.1) Determination of the h_e Parameters

(8.1.1) Common-Emitter h_e Matrix.

The common-emitter h_e matrix may be written in terms of the common-base internal admittance parameters, y_i, y_r, y_f and y_o, and the external parameters, C_e, C_c and $r_{bb'}$, as follows:*

$$\begin{bmatrix} h_{ie} & h_{re} \\ h_{fe} & h_{oe} \end{bmatrix} = \begin{bmatrix} r_{bb'} + \dfrac{1}{y_\Sigma} & \dfrac{y_r + y_o + j\omega C_c}{y_\Sigma} \\ -\dfrac{(y_f + y_o + j\omega C_c)}{y_\Sigma} & \dfrac{(y_i + j\omega C_e)(y_o + j\omega C_c) - y_f y_r}{y_\Sigma} \end{bmatrix} \quad . \quad . \quad . \quad . \quad (13)$$

where $y_\Sigma = y_i + y_r + y_f + y_o$.

(8.1.2) Common-Base Admittance Parameters.

Provided that the internal drift potential ΔV is greater than $4kT$, from Kroemer's theory[6] the internal common-base admittance parameters may be written as follows:[8]†

$$\left.\begin{aligned} y_i &\simeq \frac{G(1+X)}{2} & y_r &\simeq -\frac{GX}{K}\varepsilon^{-\psi} \\ y_f &\simeq -GX\varepsilon^{-(\psi/X)(X-1)} & y_o &\simeq \frac{G}{2K}(X-1)\varepsilon^{\psi/X} \end{aligned}\right\} \quad . \quad (14)$$

where K is the collector space-charge-widening factor, $\psi/X = \Delta V/2kT$ and $X = (1 + j8z)^{1/2}$. Here $z = (\omega w_0^2/2D_p)(kT/\Delta V)^2$. Because K is a large number, it may be seen that y_r and y_0 can be approximated to zero. Except for large values of I_e (and hence of G), for which low-level theory is in any case inapplicable, $|y_0|$ may be taken as very much smaller than ωC_c. For a first-order correction, however, any internal capacitance contributing to y_0 may be simply added to C_c.

To make the problem tractable, it remains to express y_i and y_f in terms of the dimensionless parameter $\gamma = \omega/\omega_{\alpha d}$. It has been shown[8] that $\omega_{\alpha d}$ may be closely approximated to $(\Delta V/kT)(2D_p/w_0^2)$, so that we may write $z = \gamma(kT/\Delta V)$. For γ less than 0·5 (i.e. ω less than $\omega_{\alpha d}/2$) and $\Delta V/kT$ greater than 4, the resulting z will be of the order of 0·1 or less, so that the following approximate expressions for y_i/G and $-y_f/G$ arise from eqn. (14):

$$\left.\begin{aligned} y_i/G &\simeq 1 + j2z \\ -y_f/G &\simeq (1 - 4\gamma z)(1 + j4z)\varepsilon^{-j2\gamma} \end{aligned}\right\} \quad . \quad . \quad (15)$$

Hence the real and imaginary parts of y_Σ/G are given by

$$\left.\begin{aligned} \mathscr{R}(y_\Sigma/G) &\simeq 1 - (1 - 4\gamma z)(\cos 2\gamma + 4z \sin 2\gamma) \\ \mathscr{I}(y_\Sigma/G) &\simeq 2z - (1 - 4\gamma z)(4z \cos 2\gamma \\ &\qquad - \sin 2\gamma) + \omega(C_e + C_c)/G \end{aligned}\right\} \quad (16)$$

For values of γ up to 0·5 it may be seen that $\mathscr{R}(y_\Sigma/G)$ is considerably less than $\mathscr{I}(y_\Sigma/G)$. In the subsequent analysis y_Σ/G may therefore be identified with $j\mathscr{I}(y_\Sigma/G)$.

(8.1.3) Common-Emitter h_e parameters.

In the derivation of first-order expressions for the common-emitter h_e parameters, terms involving z may be neglected, and the approximation $\varepsilon^{-j2\gamma} \simeq 1 - 2\gamma^2 - j2\gamma$ will be used where appropriate. From eqns. (13)–(16) we then have:

$$h_{ie} = r_{bb'} + \frac{1}{y_\Sigma} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17a)$$

$$h_{re} \simeq \frac{j\omega C_c/G}{y_\Sigma/G} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17b)$$

$$h_{fe} \simeq \frac{(1 - 2\gamma^2) - j(2\gamma + \omega C_c/G)}{y_\Sigma/G} \quad . \quad . \quad (17c)$$

$$h_{oe} \simeq \frac{j\omega C_c(1 + j\omega C_e/G)}{y_\Sigma/G} \quad . \quad . \quad . \quad . \quad . \quad (17d)$$

* Emitter and collector lead resistances and carrier transit time across the collector depletion-layer are here ignored.
† The emitter efficiency is assumed to be unity and recombination in the base is ignored.

621.376.3:621.396.611:654.19

The Institution of Electrical Enginee[
Paper No. 2946
July 195[

A FREQUENCY MODULATOR FOR BROADCASTING TRANSMITTERS UTILIZIN[G] OVERALL NEGATIVE FEEDBACK

By E. L. C. WHITE, M.A., Ph.D., Member.

(*The paper was first received* 18*th December*, 1958, *and in revised form* 20*th February*, 1959.)

SUMMARY

The features of existing frequency-modulation techniques are reviewed and compared with those of the feedback method. A practical embodiment of the latter is described, and is considered in some detail.

It has very few controls requiring adjustment to secure good performance, either on installation or after valve replacements, and should be particularly suited for use in unattended transmitters.

(1) INTRODUCTION

There are several well-known techniques for modulating the frequency of a carrier wave, but in order to obtain the high quality of performance required for audio broadcasting purposes, especially with regard to a.f. harmonic distortion and noise, it is usually necessary to provide a number of controls requiring careful adjustment. This arises mainly because there is no method of frequency modulation which is fundamentally linear over a considerable range of amplitude of the modulating signal.

On the other hand, there are methods of demodulation which are fundamentally linear, and the method of modulation described here uses a simple modulator having few controls and no great intrinsic linearity, the output from which is monitored by a linear demodulator whose a.f. output is applied to the modulator as negative feedback. Owing to the resulting tight control of mean frequency as well as deviation, it is possible to generate the f.m. signal directly at the desired carrier frequency, at least for the v.h.f. band II.

(2) EXISTING TECHNIQUES

Two broad classes of system have been used hitherto, namely phase modulation by the audio signal after integration, and direct frequency modulation by including as part of the oscillator frequency-determining elements an electronically variable capacitor or inductor, as in the so-called 'reactance valve'.

(2.1) Phase Modulation

The earliest f.m. stations used this technique, due to Armstrong.[1] Sinusoidal frequency modulation by a modulating angular frequency, ω_m, to give extreme angular-frequency deviations $\pm\omega_d$ is equivalent to phase modulation of extreme values $\pm\phi_d$, where $\phi_d = \omega_d/\omega_m$, which is also known as the deviation ratio or modulation index.[2] More precisely, if ϕ and ω are the instantaneous deviations, $\phi = \omega/j\omega_m$; i.e. to produce a given deviation ω by means of the phase modulation, it is first necessary to operate on the modulating signal by $1/j\omega_m$, which is integration. Clearly the greatest phase swings are required for the lowest modulating frequencies.

Armstrong[1] achieved phase modulation by adding in quadrature the result of a double-sideband suppressed-carrier amplitude-modulation process to an unmodulated carrier signal. To a first order, the resultant vector is a linearly phase-modulated carrie[r] of constant amplitude, but the linearity is poor, e.g. the har[-]monics exceed 1% for $\phi > \pm 0{\cdot}2$ rad.

Hence, in order to achieve a frequency deviation of 75 kc/[s] for a modulating frequency of 30 c/s, i.e. $\phi_d = 75\,000/30 =$ $2{\cdot}5 \times 10^3$ rad, it is necessary to use frequency multiplicatio[n] after modulation, to the extent of $2{\cdot}5 \times 10^3/0{\cdot}2 = 12\,500$. But the ratio between a typical carrier frequency of 100 Mc/[s] and the lowest frequency, e.g. 100 kc/s, at which the phas[e] modulation can be carried out without encroaching on th[e] audio spectrum, is only 1 000, so that it is necessary to inser[t] a heterodyning operation to reduce the frequency by a facto[r] of 12·5 without changing the deviation.

Altogether, assuming frequency doubling and tripling onl[y] ten stages are required before reaching the final carrier fre[-]quency. This gives plenty of opportunity for the introductio[n] of noise of all types, and considerable care is needed in th[e] design of the coupling filters to achieve phase linearity over th[e] necessary bandwidths. More recently, the 'serrasoid' system of phase modulation has somewhat reduced the difficulties b[y] making possible intrinsically linear phase modulation up t[o] nearly $\pm\pi$—an improvement on the Armstrong method by [a] factor of 15. For the example quoted above, the serrasoi[d] method would not require the heterodyne stage, and woul[d] need only seven stages to reach the final carrier. Nevertheless it is not a simple matter to achieve a low noise level.

Phase-modulation methods have the advantage that the osci[l]lator itself operates at a single frequency, so that the stability o[f] the mean carrier frequency can be simply that of a conventiona[l] crystal-controlled oscillator.

(2.2) Frequency Modulation

To achieve frequency modulation of an oscillator by mean[s] of an electronically variable reactance is basically a simpl[e] operation, but the difficulties of producing a good engineerin[g] design are associated with the requirements of linearity, stabl[e] mean frequency, and low noise level. One approach[4, 5] has bee[n] to use a balanced reactance-valve arrangement to promot[e] linearity, with a system for locking the mean phase to that of [a] crystal-controlled oscillator in order to control the mean fre[-]quency. The mean-frequency-control system necessitates fre[-]quency division of the modulated carrier by a factor sufficien[t] to reduce the maximum phase swing to less than $\pm\pi/2$, so that [a] conventional phase discriminator or 2-phase motor can be use[d]. Arising from this, it is inconvenient to modulate at a carri[er] frequency above 5 Mc/s.

Another, and rather bold, approach[6] has been to use [a] balanced reactance-valve arrangement, but to use a crystal a[s] the main frequency-determining element in the oscillator, an[d] no other mean frequency control. Since peak deviations up t[o] $\pm$125 kc/s are commonly demanded, with mean frequenc[y] stability better than $\pm$2 kc/s, this scheme requires a pair o[f] reactance valves with a differential d.c. drift less than one-sixtie[th] of the available a.c. swing, and seems to imply some selection o[f]

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. White is with E.M.I. Electronics, Ltd.

valves. The special requirements of the crystal again favour an oscillator carrier frequency below 5 Mc/s.

In both these arrangements, some care is required in adjusting the contributions from the reactance valves to the tank circuit (*LC* or crystal respectively) to be at $\pm 90°$ relative to that from the oscillation-maintaining valve, since departures from the correct phase introduce harmonic distortion in the modulation. Such adjustments require a tone source. Also, in both cases frequency multiplication by a factor of the order of 20 is required in order to produce the final carrier frequency, so that careful design is required if f.m. noise is to be kept down to the very low levels nowadays expected.

(3) NEGATIVE-FEEDBACK SYSTEM: GENERAL DESCRIPTION

The modulator to be described also uses direct frequency modulation by a reactance valve, but no particular attention is paid to linearity of the modulation characteristic. Instead, a portion of the output is heterodyned to a convenient frequency, amplitude limited, demodulated in an intrinsically linear 'diode pump' demodulator,[7] and used to provide audio-frequency negative feedback.[8] This type of demodulator is also known as an 'RC' or 'counter' demodulator (or 'discriminator'). Modulation may be carried out at or near the final carrier frequency, and the feedback is effective in reducing, not only non-linearity, but also f.m. noise of all kinds, including mains-frequency hum and microphony. A block diagram of the modulator is shown in Fig. 1.

The audio input is first operated on by a pre-emphasis amplifier, which includes facilities for adjusting the gain. The signal then passes to the main modulation amplifier, the output of which controls the frequency-modulated oscillator operating at half the radiated carrier frequency (i.e. at about 45 Mc/s for a band II transmitter). The oscillator drives a doubler and amplifier to give an output of about 10 watts. A heterodyne mixer is fed with a sample of the output and also with a stable frequency so related to the carrier frequency as to give an intermediate frequency of 1·6 Mc/s. This, carrying the whole of the original frequency deviation of the carrier, is fed to a limiter and demodulator, and the resulting a.f. output, after filtering out of the carrier, is applied to the modulation amplifier to complete the main negative-feedback loop.

The feedback extends down to zero frequency, thus stabilizing the mean carrier frequency as well as the deviation; but to give greater long-term stability to the mean carrier frequency an auxiliary centre-frequency feedback loop is provided. This operates on similar lines, but first takes the intermediate frequency of 1·6 Mc/s and reduces it to 200 kc/s by beating with a stable 1·8 Mc/s oscillation in a second heterodyne mixer. The 200 kc/s signal, carrying any drift of the carrier frequency, is applied to a second limiter and demodulator, the latter being more sensitive than that in the main feedback loop by a factor of 50. Before feeding to the modulation amplifier to complete the auxiliary feedback loop, signals above 30 c/s are filtered out, so as not to affect the audio modulation.

A second a.f. output obtained from an additional demodulator driven by the first limiter is provided at a level suitable for aural monitoring, and also supplies a peak amplitude meter calibrated in frequency deviation.

Although for ease of description the block diagram shows a dozen units, constructionally it is convenient to use only two separate chassis. An r.f. unit contains the four blocks shown

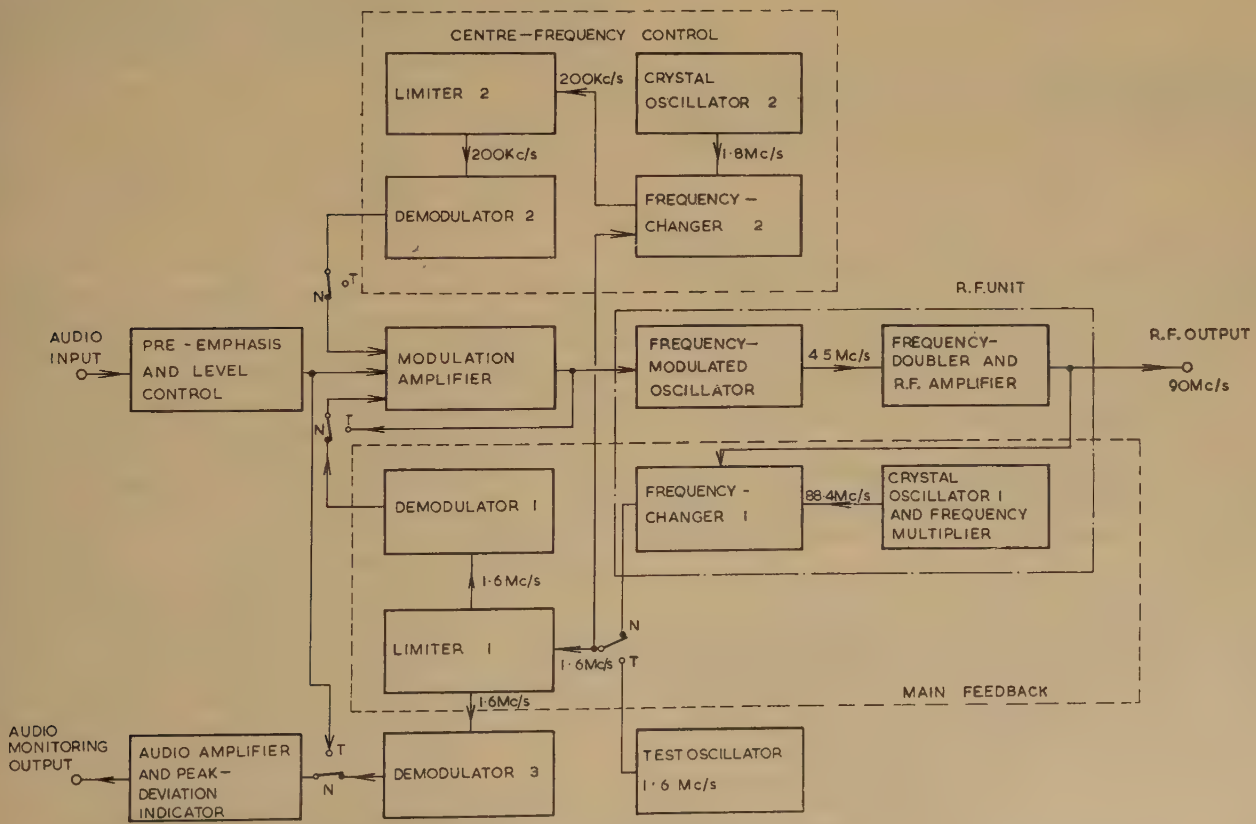


Fig. 1.—Block diagram of f.m. drive unit.
N = Normal condition of switches.
T = Test condition of switches.

inside the chain-dotted rectangle, and a main chassis contains the remainder. Two common power supplies at +300 volts and −150 volts, both stabilized, are provided from a separate power unit.

(4) DESIGN CONSIDERATIONS

(4.1) Use of 'Virtual Earth' Circuits

Considerable use is made throughout the modulator of the 'virtual earth' type of shunt feedback circuit.[9, 10] It is convenient to give a brief derivation of some useful formulae for this circuit, which is shown in generalized form in Fig. 2.

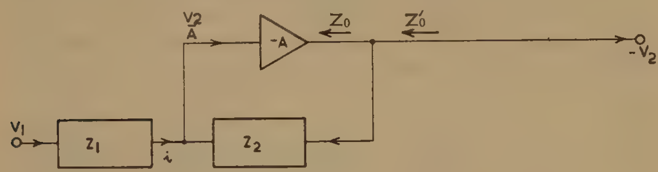


Fig. 2.—'Virtual earth' produced by shunt feedback.

The amplifier has a negative voltage gain of A, which is ideally very large. It is also assumed to have infinite input impedance, although a finite but constant impedance is easily allowed for. Let the input current be I, the feedback impedance Z_2, the output potential $-V_2$, and the input potential V_2/A. Then $IZ_2 = V_2(1 + 1/A)$, or to a first approximation when A is large, $V_2 = IZ_2$. For a given value of I or V_2 the input potential V_2/A is small for high values of A—hence the term 'virtual earth'. The impedance at the input to the amplifier, with feedback connected, is

$$\frac{V_2}{AI} = \frac{Z_2}{A\left(1 + \frac{1}{A}\right)} = \frac{Z_2}{1 + A} \quad . \quad . \quad . \quad . \quad (1)$$

If the output impedance of the amplifier without feedback is Z_0, its value with feedback applied is $Z_0' = Z_0/(1 + A)$, although this is modified if the input current is supplied through a finite impedance Z_1, becoming

$$Z_0' = \frac{Z_0}{1 + A}\left(\frac{Z_1 + Z_2}{Z_1}\right). \quad . \quad . \quad . \quad . \quad (2)$$

The overall voltage gain with a source impedance Z_1 is

$$\frac{V_2}{V_1} = \frac{Z_2/Z_1}{\left(1 + \frac{1}{A}\right) + \frac{Z_2}{Z_1}\left(\frac{1}{A}\right)} \quad . \quad . \quad . \quad . \quad (3)$$

or Z_2/Z_1 to a first approximation.

(4.2) Limiter and Demodulator

Since the limiter and demodulator are the heart of the system, they will be considered in detail first. Fig. 3 shows the essentials of the circuit.

Limiting is performed in two stages, first by current limiting in the long-tailed pair, V_1 and V_2, and secondly by limitation of potential swing at the anode of V_2 by diodes D_1 and D_2 limiting respectively at potentials V_0 and zero. C_1 provides d.c. blocking.

The resulting trapezoidal waveform of closely defined potential excursion pumps charge by means of capacitor C_2 and diodes D_3 and D_4 out of capacitor C_3 to earth. If the frequency is f, the mean current flow is basically fV_0C_2, giving the desired linear characteristic. As will be shown later, to maintain linearity the output potential swing at modulation frequencies must be small compared with V_0, and this requirement is met by using a virtual-earth amplifier with a shunt feedback resistor,

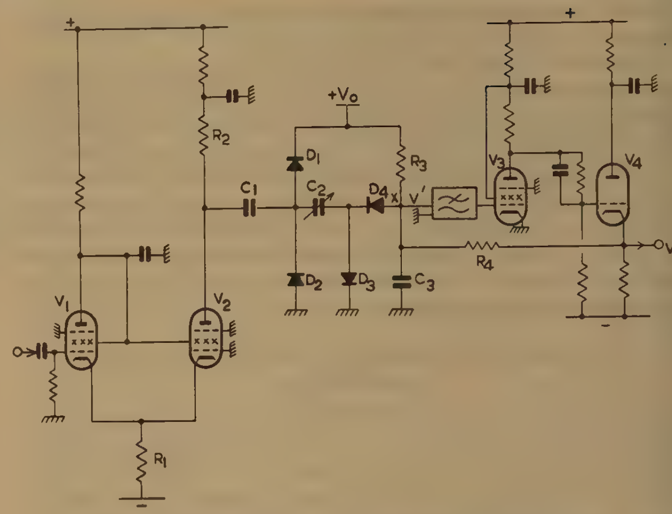


Fig. 3.—Basic circuit of limiter and demodulator.

R_4, so that the output potential swing for a deviation Δf i[s] approximately $\Delta f V_0 C_2 R_4$. The steady current $f_0 V_0 C_2$ due t[o] the carrier frequency f_0 is balanced by a resistor R_3, also supplie[d] by V_0, so that $1/R_3 = f_0 C_2$. This arrangement has the advan[-]tage that unwanted variations of V_0 affect the sensitivity t[o] frequency modulation but do not affect the carrier balance.

(4.2.1) Limiter.

Fig. 4(a) shows the input waveform at the grid of V_1. Th horizontal broken lines indicate the limiting potentials; abov the upper one V_2 is completely cut off, and below the lower on V_1 is completely cut off and V_2 carries a constant current define by the common cathode resistor R_1 and the negative suppl

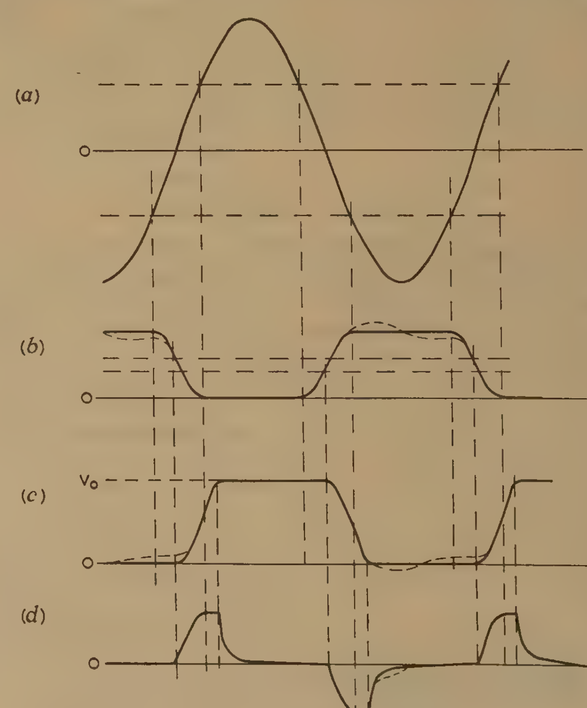


Fig. 4.—Limiter waveforms at points on Fig. 3.
(a) Grid potential of V_1.
(b) Anode current of V_2.
(c) Potential at junction of D_1 and D_2.
(d) Current through C_2.

potential. Fig. 4(*b*) shows the resulting anode-current waveform in V_2.

Since this current swing is perfectly defined, it can in principle be converted to a defined voltage swing for driving the diode pump circuit by having an appropriately low anode load resistance R_2, and this method has been used previously.[11] It is more economical in current, however, to make R_2 large, thus utilizing substantially all the available current for charging C_2 and the stray capacitances, and limiting the voltage swing at V_0 and zero respectively by the diodes D_1 and D_2;[12] the resulting voltage waveform at the junction of D_1 and D_2 is shown in Fig. 4(*c*), and the current waveform through C_2 in Fig. 4(*d*).

The anode currents of V_2 required just to reach the potential limits set by D_1 and D_2 are shown respectively by the lower and upper horizontal broken lines in Fig. 4(*b*). The vertical broken lines in Fig. 4 relate the timings of events on the various waveforms.

The extra stage of limiting is also useful in reducing the effects of grid-cathode capacitance in V_1. The current swing in V_2 is only perfectly defined for vanishingly low frequencies; at the 1·6 Mc/s carrier frequency used a capacitive current flows from the grid to the cathode of V_1 while V_1 is cut off, and thence through V_2, since the mutual conductance of the latter when conducting is much greater than $1/R_1$. This effect can add a component to the current waveform of V_2, shown dotted in Fig. 4(*b*), which is proportional to the input amplitude on the grid of V_1, and may be of the order of 5% of the ideal defined current. With no further limiting, a typical 10% unwanted amplitude modulation of the input could cause 0·5% variation of the d.c. output from the demodulator due to the carrier, or 5% of the output due to a 75 kc/s deviation signal.

The most usual cause of amplitude modulation is the effect on the frequency-modulated signal of circuits of too great a selectivity. When such circuits are exactly tuned to the carrier frequency, a second-harmonic amplitude-modulation component appears in the demodulated output. Since this output is, in this system, fed back to control the modulator performance, distortion in the demodulator causes distortion in the main f.m. output.

The diode limiters D_1 and D_2, with slope resistances of, say, 75 ohms, reduce the effect to about $\frac{1}{4}$%.

(4.2.2) Demodulators: General.

The current waveform through C_2 (Fig. 3) consists of alternate positive and negative pulses [Fig. 4(*d*)], the positive ones passing to earth through diode D_3 and the negative ones passing through D_4 to the capacitor C_3, which forms the first element in a filter for removing the carrier. It is necessary to ensure that D_3 and D_4 cannot conduct simultaneously, and with the polarities chosen, this requirement is met if the amplifier V_3 works with a few volts of negative bias. In all the applications of this type of demodulator in the equipment described, the signal excursions of grid potential of V_3 are kept to a low value. For example, Fig. 3 shows the basic circuit of demodulators 2 and 3 of Fig. 1, and in these a shunt feedback resistor R_4 is fed directly from the output of amplifier V_3, V_4.

Suppose the grid potential of V_3 is $-V'$. When the anode potential of V_2 starts to reverse its travel, there is 'lost motion' V' before C_2 starts to pass current. Thus the mean current pumped out of C_3 is

$$I = f(V_0 - V')C_2 \quad . \quad . \quad . \quad . \quad . \quad (4)$$

If I_0 and V_0' are the values of I and V' due to the unmodulated carrier of frequency f_0,

$$I_0 = f_0(V_0 - V_0')C_2 \quad . \quad . \quad . \quad . \quad . \quad (4a)$$

Differentiating eqn. (4) with respect to f gives

$$\frac{dI}{df} = (V_0 - V')C_2 \quad . \quad . \quad . \quad . \quad . \quad (5)$$

In a practical case there will be a low but finite load resistance presented at the point X (junction of D_4 and C_3 in Fig. 3) of, say, R_i. For example, demodulators 2 and 3 of Fig. 1 work into virtual-earth amplifiers as shown in Fig. 3, having an input impedance [see eqn. (1)] of $R_i = R_4/(1 + A)$. Then

$$\frac{\Delta I}{\Delta V'} = \frac{I - I_0}{V' - V_0'} = \frac{1}{R_i} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

or

$$V' = V_0' + R_i(I - I_0) \quad . \quad . \quad . \quad . \quad . \quad (6a)$$

By eqns. (5) and (6*a*),

$$\frac{dI}{df} = \left(V_0 - V_0' - R_i\Delta f\frac{dI}{df}\right)C_2 \quad . \quad . \quad . \quad (7)$$

Thus the sensitivity dI/df is not entirely independent of f, being approximately linearly related to Δf by

$$\frac{dI}{df} = \frac{(V_0 - V_0')C_2}{1 + C_2R_i\Delta f} \quad . \quad . \quad . \quad . \quad . \quad (7a)$$

If the aim is a second-harmonic component of $I \not> \frac{1}{8}$%, it can be shown that the slope dI/df can be allowed to change by 1% over the total range $\pm\Delta f$ used, i.e.

$$C_2R_i\Delta f \not> 0{\cdot}005 \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Expressed in terms of $\pm\Delta V'_{max}$ by eqns. (5) and (6),

$$\frac{\Delta V'_{max}}{V_0 - V'} \not> 0{\cdot}005 \quad . \quad . \quad . \quad . \quad . \quad (8a)$$

The current through R_3 is $(V_0 + V_0')/R_3$; hence the value of R_3 required to balance the output current at the carrier frequency f_0 is, by eqn. (4*a*),

$$R_3 = \frac{1}{f_0C_2}\left(\frac{V_0 + V_0'}{V_0 - V_0'}\right) \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The value of V_0' is the bias required by V_3 to pass a definite current determined by its anode circuit, such that the cathode of V_4 (the output point) is substantially at earth potential. Thus V_0' is likely to vary up to $\pm 0{\cdot}5$ volt from one valve sample to another of the same type. The value of V_0 for demodulators 1 and 3 of Fig. 1 is 25 volts, and for demodulator 2 is 100 volts, so that variations of V_0' are effectively swamped. It is convenient to fit a close-tolerance component for R_3 and to make C_2 a preset variable in order to adjust the output potential, V_2, to zero at the carrier frequency. This has the advantage that the sensitivity of the output to frequency modulation is then fixed, being, at f_0,

$$\frac{dI}{df} = \frac{V_0 + V_0'}{R_3f_0} \quad . \quad . \quad . \quad . \quad . \quad (10)$$

from eqns. (5) and (9).

The output conductance, $1/R$, of the demodulator proper, i.e. at point X, at carrier frequency, is [differentiating eqn. (4) by V']:

$$\frac{1}{R} = \frac{1}{R_3} + \frac{dI}{dV'} = \frac{1}{R_3} + f_0C_2$$

Therefore, from eqn. (9),

$$\frac{1}{R} = \frac{1}{R_3}\left(1 + \frac{V_0 + V_0'}{V_0 - V_0'}\right) = \frac{2}{R_3}\frac{V_0}{V_0 - V_0'}$$

i.e. the output impedance $R(Z_1$ in Fig. 2) is

$$R = \frac{R_3}{2}\left(1 - \frac{V_0'}{V_0}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

(4.3) Frequency-Modulated Oscillator

It is usual,[4–6] to separate the function of the maintenance of oscillation from that of frequency control by having separate oscillator and reactance valves. The oscillator valve supplies current in phase with the r.f. potential across the tank circuit, making good the resistive losses, and the reactance valve, by virtue of a 90° phase-shifting network in its grid circuit, supplies reactive current the amplitude of which is modulated by the instantaneous audio-frequency potential. Thus the effective capacitance and/or inductance of the tank circuit is varied, with corresponding effect on the frequency.

One of the principal advantages of the negative-feedback system is that the tight control of centre frequency permits modulation to be carried out at the radiated frequency or a low sub-multiple thereof, e.g. half, with a useful reduction of f.m. noise. However, the design of 90° phase-shifting networks to operate at 45 Mc/s without too much attenuation or dissipation, and with adequate bandwidth to give stability against temperature variations, is troublesome, in view of the significant impedances of stray capacitances and inductances at this frequency. Consequently, a little-known arrangement[13, 14] has been adopted, the basic circuit being shown in Fig. 5.

The oscillator consists of two similar valves which are effectively in parallel at radio frequency, and connected in a modified Hartley circuit with electron-coupled output. The total mean current to the two valves is substantially fixed by the common cathode-feed resistor, but the ratio in which it is shared is controlled by the modulation input, which is applied to the right-hand grid only, the left-hand one being earthed at audio frequency. The latter grid is directly connected to the tank circuit for r.f. working, but the former one has a resistance of the order of 100 ohms in series, so that, in conjunction with the stray capacitance of the grid, the r.f. phase is retarded with respect to the tank circuit.

There are thus two limiting frequencies. When the instantaneous modulating potential is sufficiently negative to cut off the right-hand section, the frequency is substantially the natural frequency of the tank circuit; but when the modulating potential is sufficiently positive to raise the common cathode potential and cut off the left-hand section, the frequency adjusts itself to a lower value such that the phase advance in the tank circuit compensates for the phase lag in the operative grid circuit. At intermediate values of modulating potential the current is shared between the two valve sections, and the resultant intermediate phase shift results in an intermediate value of frequency which changes smoothly between the limiting values as the modulating potential is varied. This arrangement has the advantage of simplicity of circuit, of well-defined output (since the total feed current is substantially fixed) and of reasonable linearity in the middle of its working range, from its symmetrical nature.

The limiting frequency shift can be calculated as follows. If the oscillator frequency is f_c, the phase-shifting resistance is R and the stray capacitance co-operating with it is C, the phase shift is (for values below about 0·5 rad) about $2\pi f_c RC$. If the

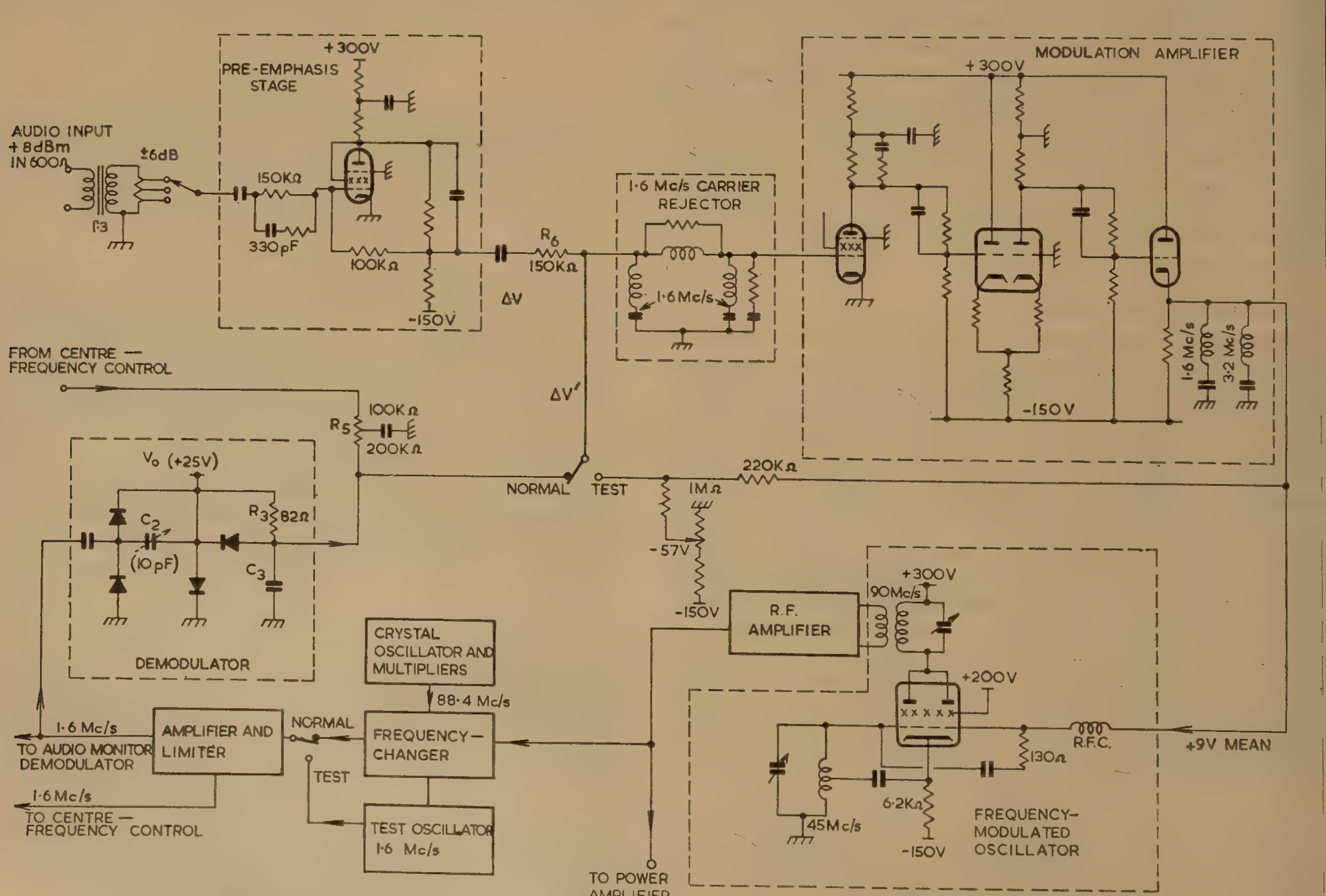


Fig. 5.—Main feedback loop.

Q-factor of the tank circuit is Q, the frequency change to give a compensating phase shift $2\pi f_c RC$ is

$$\Delta f_c = \frac{f_c(2\pi f_c RC)}{2Q} = \frac{f_c^2 \pi RC}{Q} \quad . \quad . \quad . \quad . \quad (12)$$

In calculating Q, allowance must be made for transit-time grid damping and for the damping effect of R, as well as other losses. Practical values are $R = 130$ ohms, $C = 12$ pF, $f_c = 45$ Mc/s and $Q = 33$, whence $\Delta f_c = 0{\cdot}38$ Mc/s. After frequency doubling, which reduces the risk of any maladjustment of subsequent stages affecting the oscillator, the limiting frequency swing is 0·75 Mc/s.

The modulating potential required is substantially that required to change from one section cut-off to the opposite condition in the absence of r.f. potentials, since the latter are added at nominally equal amplitudes to both grids. The middle two-thirds of the limiting frequency swing is reasonably linear (Fig. 6), and requires a modulation potential swing of

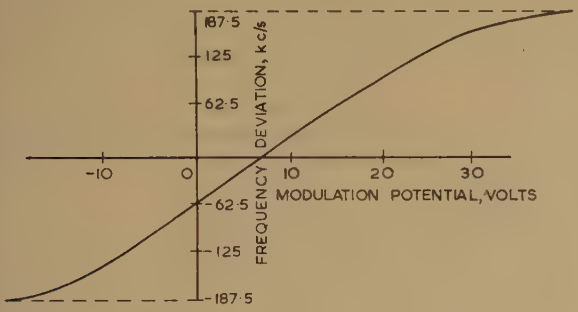


Fig. 6.—Typical characteristic of f.m. oscillator in Fig. 5.

about 30 volts for the valve used (QQVO3–10). Half the linear swing, i.e. ±125 kc/s, is regarded as available for signal modulation, leaving an equal amount available for compensating for centre-frequency drift.

(4.4) Main Feedback Loop

The essential portions of the circuit are shown in Fig. 5. The output from the pre-emphasis stage, of overall swing ΔV, is fed through a 150-kilohm resistor R_6 to the input of a 1·6 Mc/s carrier rejector filter, at which point it is combined with the outputs from the main feedback demodulator and from the centre-frequency loop. The output resistance, R, of the main demodulator, from eqn. (11), with $R_3 = 82$ kilohms, $V_0' = 3$ volts, $V_0 = 25$ volts, is 36 kilohms. The output resistance of the centre-frequency loop is 200 kilohms, hence the combined resistance at the input to the filter is 150 kilohms in parallel with 36 and 200 kilohms, i.e. 25 kilohms. The d.c. output of the main demodulator, from eqns. (4*a*) and (9) is 0·34 mA.

The output of the filter feeds the modulation amplifier, a d.c.-coupled amplifier with two stages of gain and a cathode-follower output stage. The voltage gain of the amplifier is 300 from 0 to 10 kc/s, falling thereafter to 30 at 100 kc/s. By short-circuiting the resistors between the cathodes of the second stage the gain is doubled, and this is used as a check on the loop stability.

(4.4.1) D.C. and A.F. Loop Gain.

The modulation swing required at the oscillator for 150 kc/s total deviation (after frequency doubling) is 9 volts, so that the input swing to the amplifier is 9 volts/300 = 30 mV. But in the absence of feedback the demodulator output current, from eqn. (10), is

$$\Delta I = \Delta f \frac{V_0 + V_0'}{R_3 f_0}$$

$$= \frac{150 \times 10^3 \times 28}{82 \times 10^3 \times 1{\cdot}6 \times 10^6} = 32\,\mu\text{A}$$

This, flowing into the combined resistance at the filter input of 25 kilohms (Section 4.4), gives 0·8 volt. Hence the loop gain, G, is 0·8 volt/30 mV = 26·7, or 29 dB.

It is useful to note also that the a.f. output swing, ΔV, necessary from the pre-emphasis stage is $\left(1 + \frac{1}{G}\right) \times 32\,\mu\text{A} \times 150$ kilohms, i.e. 5·0 volts.

The loop gain of 29 dB is effective in reducing harmonics due to non-linearity of the modulator by this amount, and also reduces microphony and hum and other noise similarly.

(4.4.2) Loop Gain Above Audio Frequency.

Above the audio-frequency spectrum the loop gain must fall in such a way that the system is stable, and it is desirable that the gain with feedback should have no peaks greater than 2–3 dB. For the simple type of frequency characteristic employed these objectives are satisfactorily met if the system is stable when the loop gain is raised 6 dB by the test connection provided (Section 4.4). Thus the normal loop gain when the phase has shifted 180° from the d.c. value must be less than −6 dB.

It is also necessary to attenuate the demodulator carrier frequency of 1·6 Mc/s by a large amount, since there is a considerable component of this frequency in the demodulator output, and if it reaches the modulator it will give rise to two sidebands of the radiated carrier frequency at ±1·6 Mc/s.

It is a common requirement for the level of any spurious frequency component to be −90 dB relative to the carrier, and at this low level the relative amplitude is equal to the phase modulation in radians, which can thus be $\pm 2 \times 10^{-4{\cdot}5}$ for the sum of both sidebands, or $6{\cdot}3 \times 10^{-5}$ rad. Since the modulation frequency is 1·6 Mc/s, the frequency deviation permissible is $\pm 1{\cdot}6\,\text{Mc/s} \times 6{\cdot}3 \times 10^{-5} = \pm\,100$ c/s, which is −58 dB relative to the standard ±75 kc/s deviation. The carrier output current swing of the demodulator is nearly twice the direct current, by Fourier analysis of the negative pulses only of Fig. 4(*d*), i.e. $2 \times 0{\cdot}34\,\text{mA} = 0{\cdot}68$ mA. This is 27 dB up on the ±75 kc/s current swing of 32 μA. There is also the a.f. loop gain of 29 dB to consider, so that altogether the attenuation required at 1·6 Mc/s, and ideally over a band ±75 kc/s centred on this, is 58 + 27 + 29 = 114 dB.

The solution adopted for this filtering problem is shown in Figs. 5 and 7. The total shunt capacitance in the carrier rejector filter, comprising four 200 pF capacitors, together with the 25-kilohm input resistance already mentioned, gives a 6 dB/octave cut starting at 8 kc/s. The first stage of the modulation amplifier has a step of 20 dB between 10 and 100 kc/s. The combined characteristic, shown at (*a*) in Fig. 7, reduces the loop gain to 17 dB at 15 kc/s, which still gives adequate feedback for harmonic and noise reduction, and to −6 dB at 75 kc/s, where the phase margin is 53°.

The total loss at 1·6 Mc/s is not 114 dB as required, but 66 dB, so that 48 dB remain to be found. This is done by means of the two series tuned circuits in the main filter and an additional one at the output of the modulation amplifier. The demodulator output current contains harmonics of the 1·6 Mc/s which fall off rather slowly with increasing order, as can be realized from the waveform in Fig. 4(*d*). Thus it is essential to leave some purely capacitive elements in the shunt arms of the filter. To

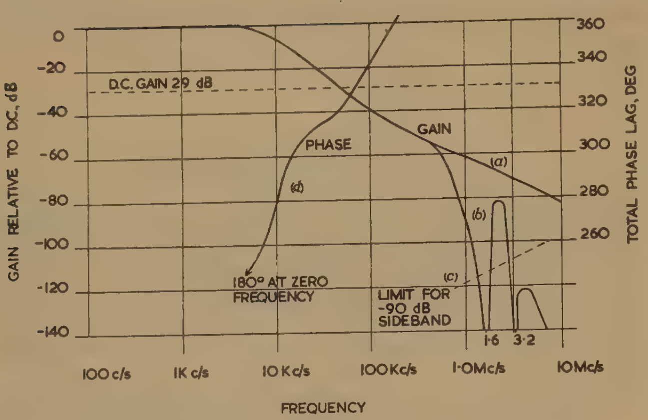


Fig. 7.—Gain and phase characteristics of main feedback loop.

ensure that the anti-resonant frequency is located midway between the fundamental and the second harmonic, these elements are equal to the capacitances in the series tuned branches. The inductance in the series arm of the filter is arranged to resonate with all the capacitances at about 0·5 Mc/s, so that it introduces very little phase shift at the −6 dB gain frequency of 75 kc/s. Damping is provided as shown to prevent any of the resonances peaking above −6 dB.

The second harmonic, 3·2 Mc/s, is still further reduced by a circuit series tuned to this frequency and connected across the modulation amplifier output. The attenuation required at the harmonic frequencies is less by 6 dB/octave, as shown at (*c*) in Fig. 7.

Curve (*b*) of Fig. 7 shows the overall characteristic, including the effects of the series tuned circuits mentioned, and also takes account of the effect of stray capacitance across the various amplifier stages, including the low-frequency equivalent of the r.f. and i.f. amplifier circuits between the frequency-modulated oscillator and the demodulator and also the mean delay of half a carrier cycle in the demodulator. The most important effect of all these is the additional phase shift they introduce at the point of −6 dB gain, which must not be allowed to exceed about 20°, i.e. a delay of 0·8 microsec. The total phase shift is shown in curve (*d*) of Fig. 7.

(4.5) Centre-Frequency Control Loop

From the frequency-changer onwards, the centre-frequency control loop (Fig. 1) is substantially the same as that shown in Fig. 3 with the addition of a capacitor across R_4 to remove a.f. modulation above 30 c/s. The object is to provide a demodulator having greater absolute stability of the frequency for zero d.c. output than that of the main feedback loop, and to feed its output back with considerably greater gain at d.c. than is provided by the main feedback. The chief problem is temperature stability. Suppose, for example, that R_3 changes by 1%; then the feedback will cause f_0 to change 1%. It is evident that, the lower the value of f_0, the less is the effect in absolute frequency shift. A value of 200 kc/s is used for f_0 in the centre-frequency control loop, because Δf must never exceed f_0, or 'negative' frequencies are produced, and 200 kc/s gives a reasonable margin of safety for a normal maximum deviation of ± 75 kc/s with occasional peaks to ± 125 kc/s.

A higher value of V_0 is practicable at 200 kc/s than at 1·6 Mc/s, and 100 volts is used, which has the advantage of reducing drift due to changing valve characteristics. A higher value of C_2 is also possible, namely 48 pF, and R_3 for balance is 100 kilohms.

By careful choice of types of component for C_2, R_3, D_1, D_2, D_3, and D_4, it has been possible to keep drift of f_0 below 1 kc/s for 50° C temperature range, and thus avoid the necessity for keeping these components in a thermostatically controlled enclosure. Silicon junction diodes are satisfactory for D_1 and D_2, and more suitable than germanium types in view of the 100 volts used for V_0 in this case. Silicon diodes are unsuitable for D_3 and D_4, because of hole storage; germanium ones have insufficiently high reverse resistance at 70° C, and therefore thermionic diodes are used. The only objection to these is that a 5% change of heater voltage gives a 300 c/s drift of f_0, but this is well within the tolerance limits.

A local feedback is provided for the virtual-earth amplifier by supplying R_4 from the output of V_4, as shown in Fig. 3. The audio frequencies are removed by shunting R_4 with a capacitor, thus reducing Z_2 (Fig. 2) at audio frequencies.

The centre-frequency feedback loop is completed by connecting the output of V_4 to the input of the modulation amplifier through R_5 (Fig. 5). The combined effect of the increased values of V_0, C_2, and the gain of V_3 (Fig. 3) is to make the d.c. gain of the centre-frequency loop 50 times that of the main loop. It follows that the capacitor across R_4, if 15 c/s modulation is not to be reduced more than 3 dB (i.e. 1 dB at 30 c/s) must have a time-constant with R_4 of at least $50/(2\pi \times 15) = 0{\cdot}5$ sec. The linearity of the demodulator in the centre-frequency loop is just as important as that of the one in the main loop, at least so far as even-order terms are concerned, since such terms not only give rise to harmonics which are unimportant here, but also to a change of d.c. output which would have the effect of changing f_0 to compensate.

The output V_2 can be indicated on a voltmeter calibrated in terms of mean frequency shift. C_2 is adjusted when an accurate 200 kc/s signal (see Section 4.7) is supplied to the limiter, to give zero potential at the cathode of V_4. Eqn. (9) gives C_2 in terms of R_3. Then, from eqns. (1), (3), (10) and (11), putting $Z_1 = R$ and $Z_2 = R_4$, the output sensitivity dV_2/df is given by

$$\frac{dV_2}{df} = \frac{V_0 + V_0'}{f_0} \times \frac{R_4}{R_3} \times \frac{1}{\left(1 + \frac{1}{A}\right) + \frac{2}{A} \times \frac{R_4}{R_3}\left(1 + \frac{V_0'}{V_0}\right)} \quad . \quad (13)$$

Provided that A is large, this is independent of valve characteristics, and it is therefore possible to pre-calibrate the meter. Absolute frequency accuracy is, however, subject to the component tolerances mentioned above, as well as—inevitably—to the accuracies of the crystal-controlled oscillators involved.

(4.6) Audio Monitor and Peak Deviation Indicator

It is convenient to drive two demodulators from the 1·6 Mc/s limiter, namely the one already discussed in the main feedback loop, and a similar one used to supply an audio monitoring output. This demodulator (No. 3 in Fig. 1) has a circuit substantially as shown in Fig. 3 from C_1 onwards, the output, V_2, from the cathode of V_4 being used for three purposes, namely

(*a*) To supply an audio output, via a buffer amplifier incorporating optional de-emphasis and an output transformer.
(*b*) To supply a peak deviation meter circuit.
(*c*) To indicate by its d.c. component on a suitably calibrated meter the drift, if any, of centre frequency, subject to the same tolerances as the main feedback demodulator.

C_2 is adjusted when an accurate 1·6 Mc/s (see Section 4.7) is supplied, to give zero potential at the cathode of V_4. Eqn. (13) with the appropriate values inserted for the circuit in question applies again, so that for function (*b*) above it is possible to pre-calibrate in terms of frequency deviation an a.f. meter circuit supplied from the cathode of V_4.

(4.7) Built-in Test Facilities

It is considered desirable to provide for two different kinds of testing, namely periodical checking of certain adjustments, and fault location. It is assumed that elementary test equipment will be available, such as a multi-range meter, tone source, absorption wavemeter, and simple cathode-ray oscillograph, and that any other essential facilities should be provided in the drive unit itself.

(4.7.1) Open-Loop Tests.

Since several negative-feedback loops of some complexity are involved, one essential facility is a means for switching these out of action or substituting more direct loops. For example, the main feedback loop, via the modulated oscillator, crystal controlled oscillator and frequency multipliers, heterodyne frequency-changer, limiter and demodulator, can be broken at the output of the demodulator and replaced by a direct feedback via a resistor from the output of the modulation amplifier to its input, as shown in Fig. 5. The resistance is chosen to give approximately the normal modulation output for a standard audio input. This test condition is essential when first tuning up all the r.f. stages, or for fault finding in them. A switch is also provided for disconnecting the feedback from the centre-frequency loop.

(4.7.2) Setting Centre Frequency.

To enable the various demodulators to be balanced by adjusting C_2 in each case, the precise centre frequency must be made available. This is done by providing a crystal-controlled oscillator operating at 1·6 Mc/s, which can be substituted for the nominal 1·6 Mc/s i.f. signal. Using the meter provided (see Section 4.6), the audio-monitor demodulator is adjusted directly, and the main-feedback-loop demodulator is then adjusted indirectly by switching to normal operation, but without the centre-frequency loop connected, and adjusting C_2 of the main demodulator until the audio-monitor demodulator d.c. output is again zero.

To adjust the 200 kc/s demodulator in the centre-frequency loop, the 1·6 Mc/s crystal is again used, and beats with the 1·8 Mc/s crystal to give an accurate 200 kc/s. Then C_2 in this demodulator is adjusted until its d.c. output is zero (see Section 4.5).

(4.7.3) Tuning.

With either the main feedback or the centre-frequency loops connected, slight detuning of the frequency-modulated oscillator will not appreciably disturb the output frequency, since the d.c. modulation output will be caused by the feedback to compensate. The correct method of final adjustment of this tuning is therefore to observe the d.c. modulation potential and tune to give it a desired value in the middle of its operating range. The remaining r.f. tuning adjustments can be carried out by normal techniques of tuning for maximum relevant valve currents or output power, followed by a final fine adjustment of one circuit (preferably the output anode circuit) to give minimum amplitude modulation when being deviated by the maximum amount of ±75 kc/s. A load to match the output impedance (e.g. 75 ohms) and a diode-probe a.m. detector are required for this purpose.

(4.7.4) Readjustment after Valve Changes.

After changing any r.f. valve, at the most it is necessary to make only two readjustments, i.e. tuning of the anode and grid circuits in an r.f. stage. In general, only the grid circuit needs adjustment. In the frequency-modulated oscillator, grid-circuit tuning as described in Section 4.7.3 is essential; but once this is done, the performance in all respects will be up to standard. All other valves can be changed without readjustment.

(5) PERFORMANCE

The representative figures given below were obtained using a quality-checking receiver which uses a 'diode pump' demodulator of the same type as those described in connection with the feedback circuits in the transmitter. Certain precautions were taken which should have still further reduced the risk of distortion in the receiver, namely two limiters in cascade were used, and a somewhat lower carrier frequency was used in the demodulator, being 1 Mc/s as against 1·6 Mc/s in the transmitter. Nevertheless, in measuring several of the characteristics it is difficult to determine how much of the departure from ideal is attributable to the measuring equipment, in particular to the receiver demodulator. It would be desirable to check the measurements with another receiver working on different principles. These remarks apply particularly to harmonic distortion and f.m. noise.

Frequency Stability.—Carrier frequency (in band II) within ±2 kc/s of nominal at any time after 10 min from switching on from cold.

Frequency Change when Modulated ±75 kc/s.—Carrier frequency changes do not exceed 300 c/s.

Modulation Frequency Response Relative to 400 c/s.—Without pre-emphasis in transmitter or de-emphasis in receiver this is not greater than ±0·1 dB from 60 c/s to 4 kc/s, ±0·5 dB from 30 c/s to 15 kc/s, and ±0·7 dB from 15 kc/s to 25 kc/s. With pre-emphasis and de-emphasis it does not exceed ±0·2 dB from 60 c/s to 4 kc/s, and ±0·6 dB from 30 c/s to 15 kc/s.

Harmonic Distortion.—The worst figure measured was 0·35% at 30 c/s with ±125 kc/s deviation. At frequencies up to 4 kc/s, with pre-emphasis and de-emphasis, the figure for ±75 kc/s varies between 0·25% and 0·15%.

Cross-Modulation.—Without pre-emphasis and de-emphasis, pairs of signals at about 4 kc/s and 10 kc/s, separated in each case by 200 c/s, were applied, each individual signal set to give 37·5 kc/s deviation. The 200 c/s beat was filtered out and measured relative to 75 kc/s deviation. At 4 kc/s it was −60 dB and at 10 kc/s it was −65 dB.

Frequency-Modulation Noise.—With pre-emphasis and de-emphasis, noise up to 15 kc/s is −70 dB relative to 75 kc/s deviation. The residue is mainly very-low-frequency components, and the C.C.I.R. psophometric weighting network increases the figure to −85 dB.

Amplitude Modulation with 75 kc/s Deviation.—At the output of the drive unit, working into a matched dummy load, the figure is 1% peak-to-peak, relative to carrier amplitude. It is sensitive to mistuning of the r.f. amplifier, and forms the best guide to the final adjustment of the output tuning.

Amplitude-Modulation Noise.—Better than −70 dB relative to carrier amplitude, the residue being mostly at mains frequency.

(6) ACKNOWLEDGMENTS

Thanks are due to Mr. D. N. Goodrum for his devoted work on the experimental side, to numerous colleagues for helpful discussions, and to the Directors of E.M.I. Electronics, Ltd., for permission to publish the paper.

(7) REFERENCES

(1) ARMSTRONG, E. H.: 'A Method of Reducing Disturbances in Radio Signalling by a System of Frequency Modulation', *Proceedings of the Institute of Radio Engineers*, 1936, **24**, p. 689.
(2) VAN DER POL, B.: 'Frequency Modulation', *ibid.*, 1930, **18**, p. 1194.
(3) DAY, J. R.: 'Serrasoid F.M. System', *Electronics*, 1948, **21**, p. 72.

(4) MORRISON, J. F.: 'A New Broadcast Transmitter Circuit Design for Frequency Modulation', *Proceedings of the Institute of Radio Engineers*, 1940, **28**, p. 444.

(5) BECK, A. C., NORBURY, F. C., and STORR-BEST, J. L.: 'Frequency Modulated V.H.F. Transmitter Techniques', *Proceedings I.E.E.*, Paper No. 2267 R, November, 1956 (**104** B, p. 225).

(6) MORTLEY, W. S.: 'Frequency-Modulated Quartz Oscillators for Broadcasting Equipment', *ibid.*, Paper No. 2302 R, December, 1956 (**104** B, p. 239).

(7) WHITE, E. L. C.: British Patent No. 741731, December, 1935.

(8) WHITE, E. L. C.: British Patent Application No. 17640, 1957.

(9) WILLIAMS, F. C.: 'Introduction to Circuit Techniques for Radiolocation', *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 303.

(10) BAXANDALL, P. J.: 'Negative Feedback Tone Control', *Wireless World*, 1952, **58**, No. 10, p. 402.

(11) SCROGGIE, M. G.: 'Low Distortion F.M. Discriminator', *ibid.*, 1956, **62**, No. 4, p. 158.

(12) WHITE, E. L. C.: British Patent No. 596459, June, 1945.

(13) Marconi's Wireless Telegraph Co. Ltd., British Patent No. 556852, October, 1942.

(14) WHITE, E. L. C.: British Patent Application 39584, 1957.

620.1.08: 621.317.432

The Institution of Electrical Engineers
Paper No. 2945 M
July 1959

THE EFFECT OF SAMPLE MOVEMENT IN FAULT DETECTION USING EDDY CURRENTS

By P. GRANEAU, B.Sc.(Eng.), Associate Member.

(*The paper was first received* 3rd *December*, 1958, *and in revised form* 20th *February*, 1959.)

SUMMARY

Eddy-current methods of non-destructive testing applied to travelling metal objects on production lines show two advantages over other techniques. These are the comparative ease with which an instrument can be coupled to test objects and the high speed of response, making possible the fast scanning rates required with fast manufacturing process.

Experience in recent years has shown that relative motion between test object and measuring head gives results which differ from those of a static system, and may severely limit permissible scanning rates. The paper draws attention to dynamically-induced currents, as causing the speed effect experienced in eddy-current testing. The theoretical arguments advanced have been used to evolve three methods of overcoming the speed effect.

LIST OF SYMBOLS

$P(z, r, \theta)$ = Cylindrical co-ordinates of a point P.
H = Magnetizing force, AT/m.
H_z, H_r = Axial and radial component of H, respectively, AT/m.
Φ_z = Axial magnetic flux, webers.
μ_r = Relative permeability.
ρ = Resistivity, ohm-m.
σ = Conductivity, mho/m.
v = Induced e.m.f., volts.
v_s = Statically induced e.m.f., volts.
v_d = Dynamically induced e.m.f., volts.
I = Primary-turn current, amp.
i = Circulating current per unit length of tube, amp/m.
f = Test frequency, c/s.
u = Linear velocity, m/s.
t = Time, sec.
a = Primary-turn radius, m.
b = Tube radius, m.

(1) INTRODUCTION

Eddy-current tests are based on self- or mutual-inductance bridges. The magnitude and phase of the bridge unbalance voltage have to be interpreted in terms of physical properties of the test object. It is therefore important to know whether and to what extent the unbalance voltage, or test signal, depends on factors not connected with material properties.

It has been found[4, 6] that relative motion between inspection coils and test object influences the test signal. This 'speed effect' can, under certain conditions, be overcome, provided that it is correctly analysed. The following Sections give an original explanation of the causes of the speed effect and the nature of the resulting signal.

(2) DYNAMICALLY INDUCED CURRENTS

Typical eddy-current tests rely on the use of coils surrounding cylindrical test specimens, e.g. tubes, rods or wires. The function of the coils is to generate an alternating electromagnetic field in which the test piece is immersed, and to sense the distribution of the resulting currents flowing in the metal. This can best be achieved by a self- or mutual-inductance bridge as shown in Figs. 1 and 2, respectively.

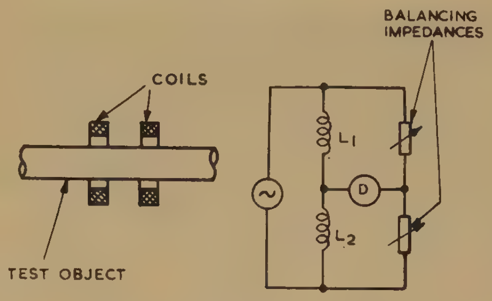


Fig. 1.—Self-inductance bridge.

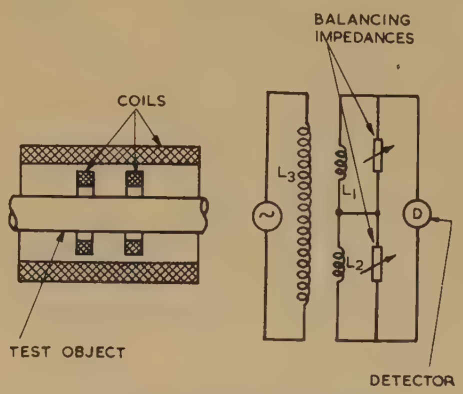


Fig. 2.—Mutual-inductance bridge.

If, with an initially balanced bridge, a long test piece is moved axially in the coil system, an unbalance voltage will occur at the bridge detector, so long as the motion persists. This voltage reverses phase with a reversal of the direction of motion, apparently indicating a current drag in the test object. The phenomenon, which can be observed at speeds as low as a few feet per minute, has variously been described as speed or current drag effect.

A convenient model showing that the drag currents are dynamically induced consists of a filamentary single circular turn, coaxial with a thin circular tube of finite but low conductivity and unity relative permeability, as shown in Fig. 3. Consider the case when the tube is removed and the wire carries a sinusoidally varying current of fixed amplitude, frequency and phase relationship, with respect to a datum time ($t = 0$).

If the single turn of radius a is centred at the origin of a

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Graneau is with British Insulated Callender's Cables, Ltd.

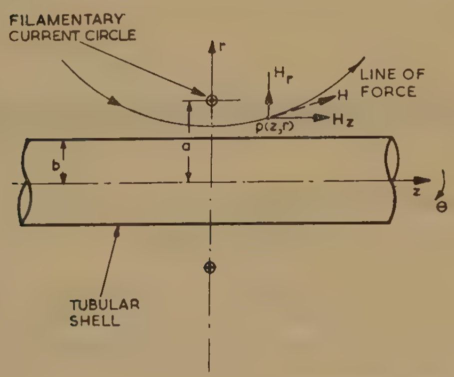


Fig. 3.—Circular thin tube coaxial with filamentary current circle.
$\mu_r = 1$, $\sigma \to 0$

cylindrical co-ordinate system and lies in the plane $z = 0$, Maxwell[8] showed that the magnetizing force, H, at any point $P(z, r, \theta)$ is a function of the solid angle subtended by the turn at P. Because of axial symmetry, H is determined by z and r only, which allows the drawing of 2-dimensional field plots. Maxwell arrives at two equations giving the axial component H_z and the radial component H_r. Both involve elliptic integrals or an infinite series of spherical harmonics.

If the tube is now inserted into the turn, circulating currents will flow in the tube wall. These modify the magnetic field and change the effective turn impedance. A measurement of the latter therefore contains information about the tube. It is common practice to interpret eddy-current measurements in terms of effective coil-impedance changes[1, 2]—a procedure which has led to considerable advances in instrument design and performance, but nevertheless affords little insight into the physical mechanism. It empirically relates observed impedance changes to observed changes in the test object. Progress probably depends on a study of the current flow and its spatial distribution in the test piece.

In a stationary thin tube all currents are circumferential, and generated only by the axial component of the field. If the tube is in motion along its axis, the radial field generates tangentially-directed e.m.f.'s, which, unless balanced by similar components on the other side of the central plane, alter the effective turn impedance. Since the radial field changes direction in the plane $z = 0$, the total circulating current on one side of the turn is decreased, whereas on the other side it is increased.

A clearer picture of the current distribution can be obtained graphically, still assuming a thin non-magnetic low-conductivity tube (Fig. 4). The insertion of the tube into the turn will only

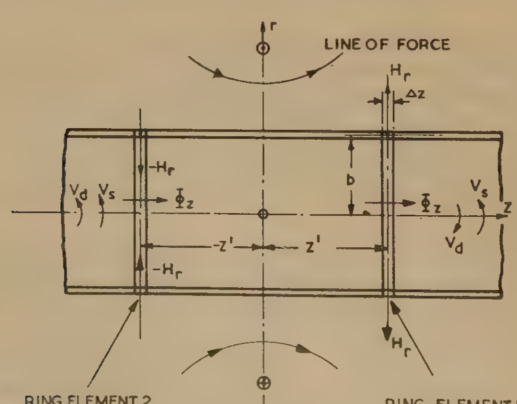


Fig. 4.—Orientation of statically and dynamically induced e.m.f.'s.

slightly change the field configuration, and Maxwell's equations for H_z and H_r are still a good approximation. The currents will be very small, but their distribution with respect to the plane of the primary turn illustrates the speed effect for an extreme case.

The current in a ring element (of width Δz) of the tube is governed by the rate of change of axial flux linking it and the rate at which the ring circumference cuts H_r. The highly resistive tube material ensures that mutually induced currents between adjacent ring elements can be neglected—a fact already implied by assuming that the currents cannot appreciably distort the original field. Two e.m.f.'s will act in ring element 1 with its centre at $(z', 0)$ as shown in Fig. 4. One is statically induced and given by

$$v_s = -2\pi f j \Phi_z \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

while the other is dynamically induced and takes the form of

$$v_d = 2\pi b u H_r \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where u is the velocity of the tube in the positive z-direction.

In the direction of tube travel, at an instant when the primary current flows clockwise, v_s acts anti-clockwise and v_d clockwise. Hence tube motion tends to reduce the current in the ring element with its centre at $(z', 0)$. In a similar ring element 2 with its centre at $(-z', 0)$ v_d will also act anti-clockwise, and the total ring current increases with tube velocity.

(3) PERIODIC DISTORTION OF CURRENT DISTRIBUTION

The equations for H_z and H_r can be expressed in different forms. Dwight[9] quotes a pair of infinite-series expressions which show rapid convergence, provided that the points (z, r) are not far removed from the z-axis. The important area, over which the field has to be explored, covers approximately a rectangle of height $2a$, length $4a$ and centre at the origin. For points inside this rectangle, the convergence of Dwight's expressions generally requires the evaluation of not more than the first four terms.

Dwight's equations for H_z and H_r are therefore a convenient starting-point for the derivation of v as a function of z, which is given in Section 8. By expressing the tube radius and z in fractions of the turn radius ($b = ka$, $z = na$), it is possible to save much arithmetic when considering a number of different coil designs. The normalized expression for H_z can easily be integrated with respect to r, allowing determination of Φ_z. Finally it is shown in Section 8 that the induced voltage in a ring element can be written as

$$v = v_d + v_s = C_1 F_{r,\,max} \cos 2\pi ft + C_2 F_{z,\,max} \sin 2\pi ft \quad . \quad (3)$$

where C_1 and C_2 are constants of a given test.

The functions F_r and F_z are sinusoidally varying quantities of amplitude $F_{r,max}$ and $F_{z,max}$ respectively. Their numerical values have been computed from eqns. (12) and (13). Fig. 5 shows the variation of $F_{r,max}$ and $F_{z,max}$ with z, where z is expressed in multiples of the turn radius a. The graph refers to a turn-tube geometry defined by $k = b/a = 0{\cdot}5$. $F_{r,\,max}$ can be positive or negative and both functions become numerically small when $z > 2a$. At large values of z, $F_{r,\,max}$ is more significant than $F_{z,\,max}$.

The function F_r generates the dynamically induced voltage v_d and will be in phase or anti-phase with the turn current, depending on the sign of $F_{r,max}$. F_z produces the statically induced voltage v_s, which lags the turn current by $\pi/2$. The phase relationship between F_r and F_z is indicated in Fig. 6, where the third co-ordinate represents time.

It is possible to choose the frequency f and the linear velocity

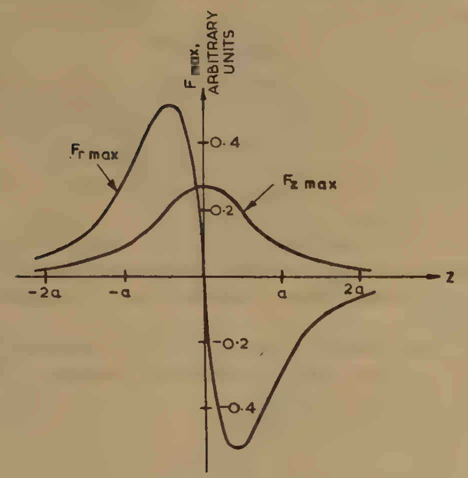


Fig. 5.—Axial and radial field component of a circular current.

$$k = \frac{b}{a} = 0{\cdot}5$$

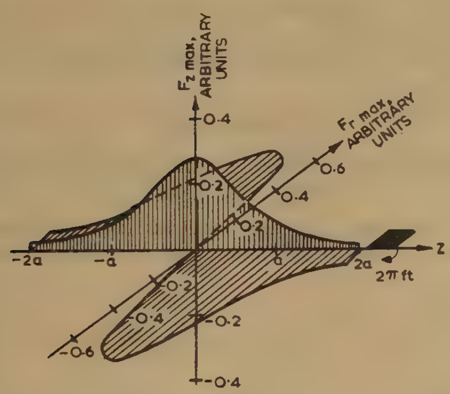


Fig. 6.—Time displacement between axial and radial component.

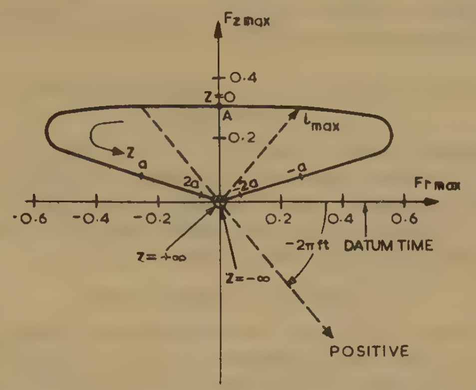


Fig. 7.—Current locus for $+\infty > z > -\infty$.

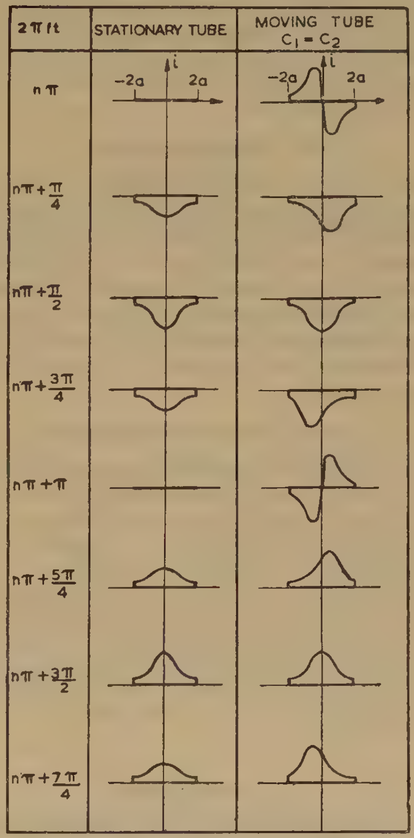


Fig. 8.—Circulating tube current distribution.

u so that $C_1 = C_2$. $F_{r,max}$ and $F_{z,max}$ in Fig. 6 will then, to another scale, be equivalent to $v_{d,max}$ and $v_{s,max}$. For any instant t' these two components can be resolved along a plane, defined by $2\pi ft'$, to give v, the total instantaneous voltage induced in a given ring element. Let the circulating current per unit length of tube have an instantaneous value i. Then i is proportional to v; and the space distribution of i, as well as the changes this distribution undergoes with time, can be determined.

At times given by $2\pi ft = n\pi$, where n is zero or a positive integer, i will have a distribution similar to that of $F_{r,max}$ in the horizontal plane of Fig. 6. The tube currents flow in opposite directions on opposite sides of the plane containing the primary turn. At instants $2\pi ft = n\pi + \pi/2$, the distribution of i is similar to that of $F_{z,max}$ in the vertical plane of Fig. 6. The turn currents then flow in the same direction on either side of the primary-turn plane.

The current distribution at any other instant can be derived from the polar locus of i_{max} for z varying between $-\infty$ and $+\infty$, as shown in Fig. 7. The area enclosed by the locus is a measure of the speed effect. For a stationary tube the locus is a straight line along the $F_{z,max}$-axis, and the enclosed area vanishes. To obtain the distribution of i at an instant t', it is necessary to project every point of the locus on a straight line through the origin and inclined at an angle $2\pi ft$ to the datum time line, which, in Fig. 7, is the positive $F_{r,max}$-axis.

It is thus possible to tabulate (Fig. 8) the distribution of i at intervals of one-eighth of a cycle for the stationary tube and for the special case $C_1 = C_2$. With a stationary tube the current is always symmetrical about the plane containing the primary turn. Movement of the tube at constant velocity distorts the current distribution. Furthermore, the vectors representing i_{max} and associated with the positive z-direction lag behind the corresponding vectors associated with the negative z-direction.

It can easily be shown that the use of two coaxial parallel turns cannot restore current symmetry. The self-inductance bridge (Fig. 1) often employed for eddy-current testing, which

approximates to the case of two turns, will therefore show a speed effect, the severity of which will increase with speed.

(4) THREE METHODS OF OVERCOMING THE SPEED EFFECT

High-speed testing is most desirable when a large volume of material has to be scanned for defects. The three practical solutions mentioned therefore deal with defect detection only, and it is assumed that the test object moves simultaneously in two adjacent arms of a bridge circuit.

(4.1) Mutual-Inductance Bridge

If the primary coil L_3 of a mutual inductance bridge, as shown in Fig. 2, is very long compared with the secondary-coil width and spacing, all dynamically induced currents near the end planes of L_3 are far removed from the pick-up coils and consequently cannot affect the bridge balance. Moreover, by keeping the currents in the secondary coils L_1 and L_2 very small, the resulting magnetic field within the effective range of the secondary coils is purely axial and independent of the speed of the test object. This arrangement is being used with great success for wire testing at speeds of up to 5 000 ft/min.

(4.2) Phase Suppression

The statically and dynamically induced voltage components act along fixed lines in space, but reach their maximum values at different instants of time. The idealized example already mentioned produces a phase displacement of 90° between the two voltages. Whatever the value of this angle in a practical example, for a given case it remains constant, and therefore that component of the bridge unbalance voltage which is due to dynamically induced currents can be suppressed by conventional phase-discriminating circuits.

(4.3) Pulse-Length Discrimination

The inertia of plant moving metallic test objects at high speed is likely to be large. Rapid fluctuations in speed are therefore not expected and the speed signal will either be constant or slowly varying. On the other hand, a defect travelling past narrowly-spaced coils gives rise to a short pulse. It is easy to devise filters which pass the defect pulses and effectively block slow changes in the signal level. The amplitude of a pulse may still depend, to some degree, on the test speed, because a defect can interfere with the flow of dynamically induced currents. This should be considered when the pulse amplitude is taken as a measure of defect size.

(5) CONCLUSION

Analysis of an idealized test system indicates that the radial field component of an encircling coil generates currents in a moving test object. Since the test signal is a function of all the currents flowing in the test piece, it will contain a component related to the relative velocity between detector head and test piece. The speed modulation of the test signal predicted by the field analysis is consistent with the observed speed effect in eddy-current testing. Three established methods of dealing with the speed effect have been briefly outlined in Section 4.

(6) ACKNOWLEDGMENTS

Acknowledgment is due to Dr. L. G. Brazier, Director of Research and Education of British Insulated Callender's Cables, Ltd., for permission to publish the paper, and to Prof. J. E. Parton of the University of Nottingham under whose supervision a study of the theory of eddy-current testing is being carried out. The author wishes to express his gratitude to Mr. S. A. Swann of the University of Nottingham for very helpful suggestions.

(7) REFERENCES

(1) HOCHSCHILD, R.: 'Eddy-Current Testing by Impedance Analysis', *Nondestructive Testing*, 1954, **12**, p. 35.

(2) HOCHSCHILD, R.: 'The Theory of Eddy-Current Testing in One (Not-so-easy) Lesson', *ibid.*, 1954, **12**, p. 31.

(3) HOCHSCHILD, R.: 'Testing Metals "On-the-Fly" with Eddy Currents', *Control Engineering*, 1957, **4**, p. 79.

(4) JOSEPH, H. M., and NEWMAN, N.: 'Eddy-Current Mutual Inductance Transducers with High-Conductivity Reference Plates', *Transactions of the American I.E.E.*, 1955, **74**, Part I, p. 39.

(5) ALLEN, J. W., and OLIVER, R. B.: 'The Inspection of Small-Diameter Tubing by Eddy-Current Methods', *Nondestructive Testing*, 1957, **15**, p. 104.

(6) ROBINSON, J. B., and PERRIAM, L. T.: 'The Inspection of Tubes for Flaws and Variations in Thickness', *Journal of the Institute of Metals*, 1958, **86**, p. 316.

(7) STANFORD, E. G., and TAYLOR, H. W.: 'Eddy Current Instruments for Use in the Field of Non-Destructive Testing', B.I.O.S. Final Report No. 1791, H.M. Stationery Office.

(8) MAXWELL, C.: 'Electricity and Magnetism', Vol. II (Clarendon Press, 1904).

(9) DWIGHT, H. B.: 'Electrical Coils and Conductors' (McGraw-Hill, 1945).

(8) APPENDIX

To Calculate the Distribution along the z-axis of Tube Voltages and Currents

Dwight[9] states that, for points not far from the z-axis, values for H_z and H_r (Fig. 3) are best computed from the following expression:

$$H_z = \frac{2\pi I a^2}{10\rho^3}\left[P_1'(\mu) - \frac{1}{2}\frac{r^2}{\rho^2}P_3'(\mu) + \frac{1\times 3}{2\times 4}\frac{r^4}{\rho^4}P_5'(\mu) - \ldots\right] \quad . \quad (4)$$

$$H_r = \frac{2\pi I a^2 r}{10\rho^4}\left[\frac{1}{2}P_2'(\mu) - \frac{1\times 3}{2\times 4}\frac{r^2}{\rho^2}P_4'(\mu) + \frac{1\times 3\times 5}{2\times 4\times 6}\frac{r^4}{\rho^4}P_6'(\mu) - \ldots\right] \quad . \quad (5)$$

where I = Current in the circular turn.

a = Radius of current circle.

$$\mu = \frac{z}{\rho}$$

$$P_n(\mu) = \frac{1}{2^n n!}\frac{\partial^n}{\partial\mu^n}(\mu^2 - 1) \text{ surface zonal harmonics.}$$

$$P_n'(\mu) = \frac{\partial}{\partial\mu}P_n(\mu)$$

$$\rho^2 = z^2 + a^2$$

For the purpose of the investigation it was convenient to make the following substitutions in Dwight's expressions:

$$z = na;\ r = ka \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Hence

$$H_z = \frac{2\pi I}{10a(n^2+1)^{3/2}}\left[P_1'(\mu) - \frac{1}{2}\frac{k^2}{n^2+1}P_3'(\mu) + \frac{1\times 3}{2\times 4}\frac{k^4}{(n^2+1)^2}P_5'(\mu) - \frac{1\times 3\times 5}{2\times 4\times 6}\frac{k^6}{(n^2+1)^3}P_7'(\mu)\right] \quad . \quad (7)$$

$$H_r = \frac{2\pi Ik}{10a(n^2+1)^2}\left[\tfrac{1}{2}P_2'(\mu) - \frac{1\times 3}{2\times 4}\frac{k^2}{n^2+1}P_4'(\mu) + \frac{1\times 3\times 5}{2\times 4\times 6}\frac{k^4}{(n^2+1)^2}P_6'(\mu) - \frac{1\times 3\times 5\times 7}{2\times 4\times 6\times 8}\frac{k^6}{(n^2+1)^3}P_8'(\mu)\right] \quad . \quad (8)$$

where $$\mu = \frac{n}{\sqrt{(n^2+1)}}$$

The expressions have, in this way, been normalized for any turn radius a. It will also be noticed that only the first four terms have been retained, which give sufficient accuracy in the region of practical importance limited by $n > 0\cdot 2$, $k < 0\cdot 8$.

v_d is proportional to H_r (u = constant), but v_s can only be derived from H_z by means of an integration leading to Φ_z.

$$\Phi_z = \int_0^b 2\pi r H_z dr \quad . \quad . \quad . \quad . \quad . \quad (9)$$

but $$r = ka,\ b = k'a,\ \frac{dr}{dk} = a$$

Therefore $$\Phi_z = \int_0^{k'} 2\pi a H_z k a dk = 2\pi a^2 \int_0^{k'} H_z k dk \quad . \quad . \quad (10)$$

Substituting for H_z and integrating term by term leads to

$$\Phi_z = \frac{4\pi^2 Iak^2}{10(n^2+1)^{3/2}}\left[\tfrac{1}{2}P_1'(\mu) - \frac{1\times 1}{4\times 2}\frac{k^2}{n^2+1}P_3'(\mu) + \frac{1\times 1\times 3}{6\times 2\times 4}\frac{k^4}{(n^2+1)^2}P_5'(\mu) - \frac{1\times 1\times 3\times 5}{8\times 2\times 4\times 6}\frac{k^6}{(n^2+1)^3}P_7'(\mu)\right] \quad . \quad . \quad . \quad . \quad (11)$$

Let a function F_z be defined by

$$F_z = \frac{k^2}{(n^2+1)^{3/2}}\left[\tfrac{1}{2}P_1'(\mu) - \frac{1\times 1}{4\times 2}\frac{k^2}{n^2+1}P_3'(\mu) + \frac{1\times 1\times 3}{6\times 2\times 4}\frac{k^4}{(n^2+1)^2}P_5'(\mu) - \frac{1\times 1\times 3\times 5}{8\times 2\times 4\times 6}\frac{k^6}{(n^2+1)^3}P_7'(\mu)\right] \quad . \quad . \quad . \quad . \quad (12)$$

and a function F_r by

$$F_r = \frac{k}{(n^2+1)^2}\left[\tfrac{1}{2}P_2'(\mu) - \frac{1\times 3}{2\times 4}\frac{k^2}{n^2+1}P_4'(\mu) + \frac{1\times 3\times 5}{2\times 4\times 6}\frac{k^4}{(n^2+1)^2}P_6'(\mu) - \frac{1\times 3\times 5\times 7}{2\times 4\times 6\times 8}\frac{k^6}{(n^2+1)^3}P_8'(\mu)\right] \quad . \quad . \quad . \quad . \quad (13)$$

It follows that

$$v_s \propto \Phi_z \propto F_z[k, n, P(\mu)]$$
$$v_d \propto H_r \propto F_r[k, n, P(\mu)]$$

Both v_d and v_s will be sinusoidally varying quantities, and their sum can be expressed as

$$v = v_d + v_s = C_1 F_{r,\,max} \cos 2\pi ft + C_2 F_{z,max} \sin 2\pi ft$$

which is, in fact, eqn. (3).

621.396.677.73:621.372.81

The Institution of Electrical Enginee
Paper No. 2874
July 19

THE EFFECT OF FLANGES ON THE RADIATION PATTERNS OF WAVEGUIDE AND SECTORAL HORNS

By P. C. BUTSON, M.Sc., and G. T. THOMPSON, B.Sc.(Eng.), Associate Member.

(*The paper was first received* 1st *May*, 1958, *and in revised form* 1st *January*, 1959.)

SUMMARY

Measurements are given of radiation patterns of open-ended rectangular waveguide with conducting flanges attached to the long edges of the aperture. In particular, the considerable effect of different flange lengths and included angles on the E-plane patterns is examined; these measurements are made at three frequencies in order to supplement data, already published, describing measurements made at a single frequency. A suggestion that the E-plane radiation pattern can be approximated by the radiation pattern of a primary source and two secondary sources near the end of the flanges has been substantiated if a suitable obliquity factor is assumed. Some E-plane patterns of flanged H-plane sectoral horns are also given. It is found that there are differences between these patterns and those of similarly flanged waveguide.

The effect of flanges on the H-plane patterns of an E-plane sectoral horn is comparatively small, as was expected.

(1) INTRODUCTION

Measurements of the E-plane radiation patterns of small flanged H-plane sectoral horns, at one frequency, have been published by Owen and Reynolds. Further information about the frequency dependence of such patterns has been obtained. For convenience, the majority of measurements were made using open-ended WG10 waveguide as the basic aperture. It was thought likely that the results of these measurements would provide useful information for the design of flanged sectoral horns. This was found to be so, but the limited results taken with sectoral horns indicate that there are differences which become more marked for longer flanges.

The effect of flanges on the patterns of E-plane sectoral horns was also investigated. Although the modification of the patterns by the flanges was expected to be relatively small, the use of flanges is sometimes desirable for mechanical reasons.

(2) METHOD OF MEASUREMENT

The patterns were plotted automatically. The aerial under test was rotated about a vertical axis through its aperture, together with a crystal detector which fed the recording amplifier through a flexible cable. The recording chart was driven in synchronism with the rotation of the aerial.

Only small deviations from symmetry were present in the measured patterns. The chart record obtained was folded through the centre and a new curve drawn whose ordinates were the average of the ordinates on either side of the original pattern. A symmetrical radiation pattern was thus produced, having smaller errors than the original.

The overall amplitude response of the system was determined and was used to produce corrected power response patterns normalized to unit power.

The waveguide aperture measurements were made at a range of 12·5 ft and the sectoral horn measurements at a range of 66 ft.

The transmitting aerial was a pyramidal horn with an apertu of 8 in square.

(3) RADIATION PATTERNS OF WG10 WAVEGUIDE APERTURE WITH FLANGES ON THE LONG SIDES

(3.1) E-Plane Patterns

Radiation patterns were taken for seven flange lengths, L, each of five included angles, A (Fig. 1). The lengths vari from $\lambda_0/4$ to $7\lambda_0/4$, where λ_0 is the free-space wavelength

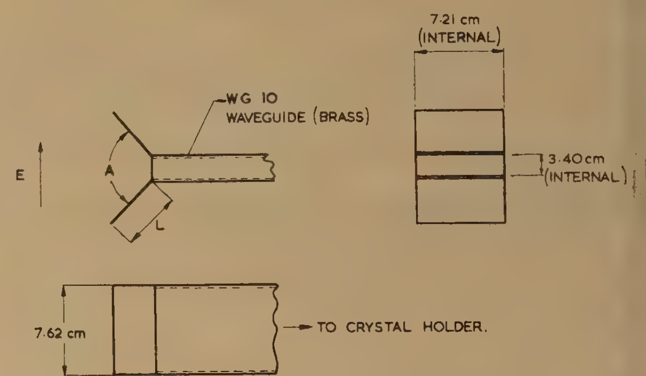


Fig. 1.—Flanged WG10 waveguide aperture.

1·52 times the cut-off frequency, f_c, of the waveguide. Tl included angles were 90°, 120°, 180°, 240°, and 300°. Eac different flange configuration was measured at three frequencie 2·59, 3·16 and 4 Gc/s, equivalent to $1{\cdot}25f_c$, $1{\cdot}52f_c$ and 1·92 respectively. The upper and lower frequencies were determine by the range of the signal source used.

The patterns obtained are shown in Fig. 2. It should l noted that these patterns cannot be compared directly with tho of Reference 1, because Owen and Reynolds used an E-pla aperture of 1·00 in, whereas WG10 waveguide has an E-pla dimension of 1·34 in.

It has been suggested that the effect of the flanges on tl radiation pattern of a waveguide aperture is similar to that secondary sources near the free ends of the flanges.[1] To confir this, some analysis of the experimentally obtained patterns h been made. The positions of the maxima on all the patter for $A = 180°$ were compared with those calculated for tw secondary sources at the ends of the flanges, and a source in tl aperture of the waveguide. The positions of the maxima we calculated only for $A = 180°$, because explicit expressions cou not be obtained for other angles, and for this value of A th are given by

$$\theta = \arc\sin \frac{2n\lambda}{B} \text{ if } \cos\phi > -2k$$

and $$\theta = \arc\sin \frac{(2n+1)\lambda}{B} \text{ if } \cos\phi < 2k$$

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

The authors are at the Research Laboratories of The General Electric Company, Limited, Wembley, England.

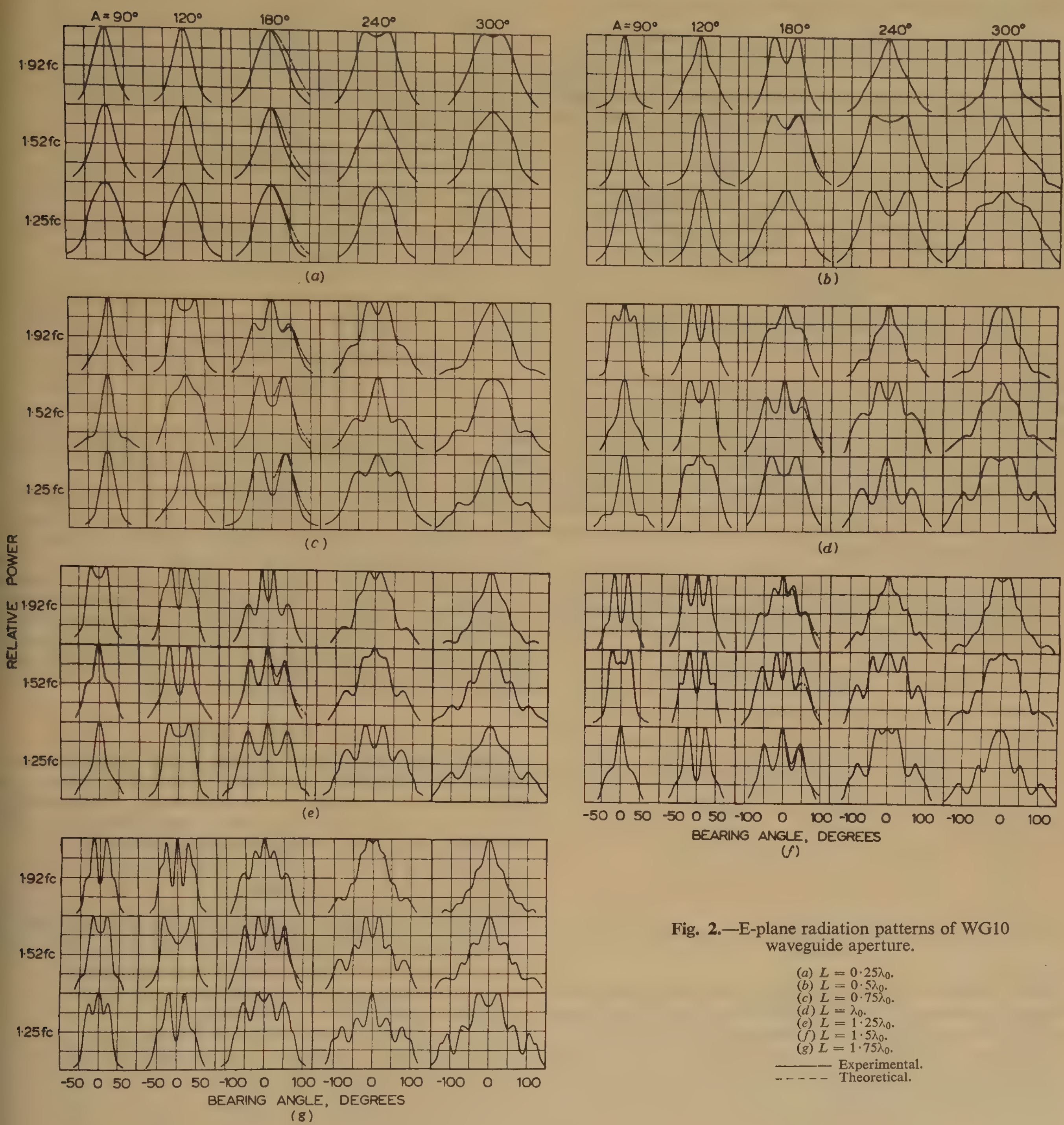


Fig. 2.—E-plane radiation patterns of WG10 waveguide aperture.

(a) $L = 0 \cdot 25\lambda_0$.
(b) $L = 0 \cdot 5\lambda_0$.
(c) $L = 0 \cdot 75\lambda_0$.
(d) $L = \lambda_0$.
(e) $L = 1 \cdot 25\lambda_0$.
(f) $L = 1 \cdot 5\lambda_0$.
(g) $L = 1 \cdot 75\lambda_0$.
——— Experimental.
- - - - - Theoretical.

where

θ = Bearing angle.
$B/2$ = Distance between the primary source and each secondary source.
k = Amplitude of each secondary source relative to the primary source.
ϕ = Phase of each secondary source relative to the primary source.

The radiation pattern produced by these three sources is given by

$$P_\theta = 1 + 4k \cos\left(\frac{\pi B}{\lambda} \sin\alpha \sin\theta\right)\left\{\cos\left[\frac{\pi B}{\lambda}\cos\alpha\cos\theta + \phi\right] + k\cos\left(\frac{\pi B}{\lambda}\sin\alpha\sin\theta\right)\right\}$$

where P_θ = square of amplitude and 2α = angle subtended at the primary source by the line joining the two secondary sources.

For collinear sources ($\alpha = \pi/2$) this expression reduces to

$$P_\theta = 1 + 4k \cos\left(\frac{\pi B}{\lambda}\sin\theta\right)\left[\cos\phi + k\cos\left(\frac{\pi B}{\lambda}\sin\theta\right)\right]$$

Fig. 3 shows the corresponding vector diagram, from which it will be seen that when $\phi = n\pi$ the variation of P_θ with θ, i.e.

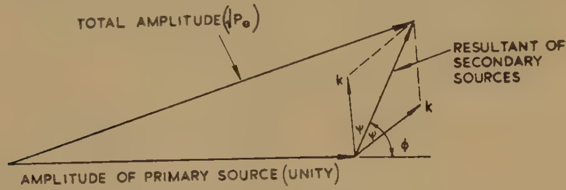


Fig. 3.—Vector diagram for three collinear sources.
$\psi = (\pi B/\lambda)\sin\theta$

the ripple on the pattern, is large, whereas when $\phi = (n + \frac{1}{2})\pi$ the variation is small. $\phi = 2n\pi$ produces a maximum on the axis of the radiation pattern, and $\phi = (2n + 1)\pi$ produces a minimum. A more detailed investigation shows that, as ϕ is changed, the variation in the amplitude of the ripple on the radiation pattern is approximately sinusoidal, so that $V_1/V_0 \simeq 2k\cos\phi$ near the axis of the pattern, where V_0 is the mean amplitude and V_1 is half the peak-to-peak amplitude of the ripple on the amplitude pattern.

The maximum and minimum values of the experimental radiation patterns near the axis were used to obtain an estimate of k. The variation of V_1/V_0 as a function of $B/2\lambda$ is shown in Fig. 4, the value being plotted as positive when the pattern has

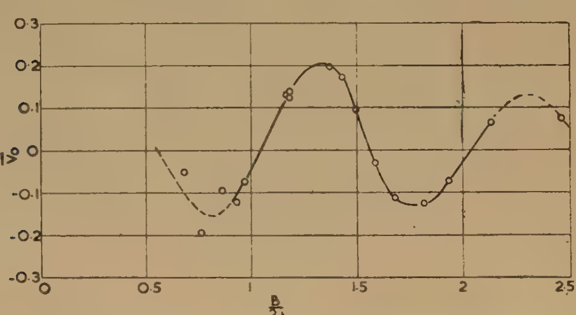


Fig. 4.—V_1/V_0 as a function of the distance between the sources for $A = 180°$.

a maximum on the axis and as negative when the pattern has a minimum on the axis. From the extreme values of the smooth curve, an average value of $k = 0{\cdot}077$ was obtained.

The phase of the secondary sources shown in Fig. 5 was obtained from Fig. 4 by plotting the maxima, minima and zeros of V_1/V_0, taking ϕ as $2n\pi$, $(2n + 1)\pi$ and $(2n \pm \frac{1}{2})\pi$ respectively, against the corresponding values of $B/2\lambda$. The fact that these points lie very nearly on a straight line of slope 2π shows that the secondary sources can be considered to be excited by energy propagated along the flange with free-space velocity. Extrapolating this graph to $\phi = 0$ shows that the auxiliary sources are in phase with the primary source when $B/2\lambda = 0{\cdot}31$. However, for very short flanges the representation by three sources appears to become less valid.

Taking $k = 0{\cdot}077$ and obtaining the appropriate value of ϕ from Fig. 5, theoretical patterns corresponding to a selection of the experimental patterns were calculated, using an overall obliquity factor of $1 + \cos\theta$. These are shown as broken

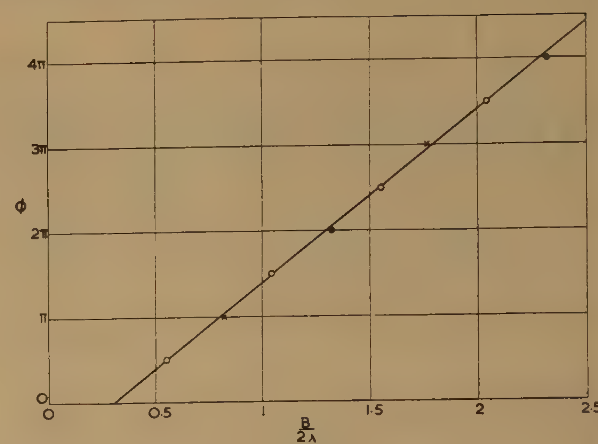


Fig. 5.—Variation of phase of the secondary sources.

○ Zeros, $\phi = n\pi + \pi/2$
× Minima, $\phi = (2n + 1)\pi$
● Maxima, $\phi = 2n\pi$

curves in Fig. 2. It is seen that there is good agreement in th positions of the maxima and minima and fair agreement in the values.

Only two patterns were calculated for values of A other tha 180°. Agreement with experiment was quite good for value of θ within the angles of the flanges. Outside this angle th calculated value was too high. This may be due to the fa that, for such values of θ, the waveguide aperture is parti screened. There is some doubt concerning the correct value o B which should be used in the calculations for flange angles othe than 180°, but the variations in the calculated patterns caused b small variations in the assumed value of B do not significantl affect the agreement with the measured results.

(3.2) H-Plane Patterns

It was not expected that the H-plane patterns would be muc affected by the presence of flanges, at least when these wer short or were in or behind the aperture plane. Measuremen made at 3·16 Gc/s, given in Table 1, confirmed this. Th half-power beam width of the unflanged aperture was 64°.

Table 1

H-Plane Half-Power Beam Widths for Flanged WG1 Waveguide Aperture at 3·16 Gc/s

(Flanges attached to long aperture dimension)

L/λ_0	Angle A				
	90°	120°	180°	240°	300°
0·25	60°	62°	65°	68°	68°
0·5	58°	57°	61°	56°	60°
0·75	60°	68°	72°	55°	60°
1·0	54°	56°	73°	64°	58°
1·25	49°	98°	63°	—	—
1·5	42°	80°	60°	—	—

(4) COMPARISON OF E-PLANE RADIATION PATTERNS OF FLANGED WAVEGUIDE AND SECTORAL HORNS

Measurements were made on an H-plane sectoral horn wi an aperture height of 18 in and a width of 1·34 in (i.e. the sam width as WG10 waveguide). The length of the horn from th aperture to the junction with the waveguide feed was 18·4 in.

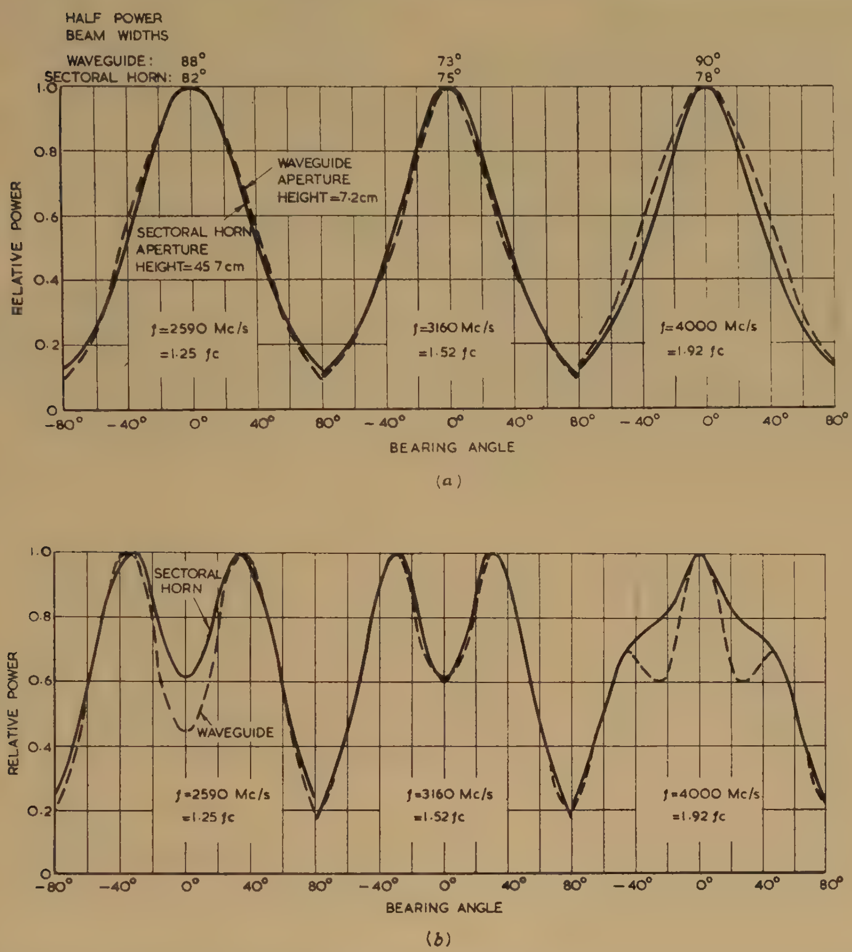


Fig. 6.—Comparison of E-plane radiation patterns of flanged waveguide and an H-plane sectoral horn of the same aperture width.

(a) $A = 180°$, $L = 0{\cdot}25\lambda_0$.
(b) $A = 180°$, $L = 0{\cdot}75\lambda_0$.

E-plane radiation patterns of this horn were measured with quarter and three-quarter wavelength flanges in the plane of the aperture (i.e. $L = 0{\cdot}25\lambda_0$ and $0{\cdot}75\lambda_0$, $A = 180°$). The patterns are shown in Fig. 6 superimposed on the corresponding WG10 waveguide patterns, from which it is seen that there are differences, particularly for the longer flanges.

(5) EFFECT OF FLANGES ON THE H-PLANE PATTERNS OF AN E-PLANE SECTORAL HORN

The effect of flanges on the long sides of an E-plane sectoral horn, having an aperture height of 12 in and an aperture width of 2·84 in, was found to be small, as may be seen from Table 2. The length of the sectoral horn to the junction with the waveguide feed was 12·4 in). The patterns are generally rather narrower than those for the unflanged horn, especially when the included angle, A, is not greater than 180°.

(6) CONCLUSIONS

It is clear that flanges normal to the electric field may be used to control the E-plane patterns of small radiating apertures, as stated by Owen and Reynolds. The effect of such flanges can be represented by secondary line sources at the free ends of the flanges, each having an amplitude of about 0·077 times that of the primary source in the waveguide aperture.

Table 2

H-PLANE HALF-POWER BEAM WIDTHS FOR FLANGED E-PLANE SECTORAL HORNS

A	L/λ_0	Beam width		
		at 2·59 Gc/s	at 3·16 Gc/s	at 4 Gc/s
deg		deg	deg	deg
—	0	70	63	58
90	0·25	64	56	53
90	0·5	58	48	43
180	0·25	65	58	57
180	0·5	64	60	57
180	1·0	69	61	59
240	0·5	68	62	58
240	1·0	77	65	58
270	0·25	68	64	60
270	0·5	70	63	60
270	0·75	74	62	58
270	1·0	76	63	58

Measurements on a flanged waveguide aperture show that such flanges have little effect on the H-plane patterns.

A limited number of measurements on apertures extended in the H-plane (H-plane sectoral horns) show that the results obtained with the flanged waveguide give patterns which are similar to those of a flanged sectoral horn, although the latter are likely to show differences in the pattern detail, especially if the flanges are long. Owen and Reynolds found that the E-plane patterns were substantially independent of aperture height, but their measurements were all made with sectoral horns.

It would therefore appear that, considering the transition from a waveguide aperture to a sectoral horn of increasing H-plane aperture, the change of radiation pattern is at first large but becomes smaller for longer apertures. It is not, however, known how the constancy, or otherwise, of the horn length affects these results, as the paper by Owen and Reynolds makes no mention of the length of the horns used, and for the present pap measurements were made on only one H-plane sectoral horn.

The effect of flanges tangential to the electric field on t H-plane patterns of E-plane sectoral horns is small, owing the weaker excitation of the flanges.

(7) ACKNOWLEDGMENT

The authors are indebted to the Admiralty for permission publish the paper.

(8) REFERENCE

(1) Owen, A. R. G., and Reynolds, L. G.: 'The Effect Flanges on the Radiation Patterns of Small Horn *Journal I.E.E.*, 1946, **93,** Part IIIA, p. 1528.

621.372: 621.317.332.2

The Institution of Electrical Engineers
Paper No. 2881 E
July 1959

THE ESTIMATION OF THE REACTANCE OF A LOSS-FREE SURFACE SUPPORTING SURFACE WAVES

By K. P. SHARMA, M.Sc., Ph.D., Graduate.

(*The paper was first received* 22*nd October*, 1958, *and in revised form* 28*th January*, 1959.)

SUMMARY

The transmission-line analogue technique has been applied to calculate the reactances of a few types of surface which may support surface waves. Experimental methods of estimating the reactances of such surfaces have been discussed, with special reference to the difficulties encountered in practice. Results obtained by theoretical calculations have been found to agree with experimental results within the limits of experimental error and the validity of the assumptions made in the theoretical calculations.

LIST OF PRINCIPAL SYMBOLS

E_x, E_y = Components of the electric field.
H_z = Magnetic field.
β, β_0 = Phase-change coefficients of a plane wave in a medium and in free space.
X_s = Surface reactance of the guiding surface.
β_s = Surface-wave phase-change coefficient.
λ_s, λ = Surface-wave wavelength and free-space wavelength of a plane wave.
u = Surface-wave vertical decay coefficient.
μ_0 = Permeability of free space.
ϵ_0, ϵ_r = Permittivity of free space and relative permittivity of a medium.
Z_0, Z = Characteristic wave impedance of free space and that of medium.
A = Amplitude factor for the surface wave field.
a, d = Slot depth and slot width of corrugation respectively.
D = Pitch of corrugation.
h = Thickness of dielectric coating.
r = Radial distance.

(1) INTRODUCTION

Surface waves[1] are electromagnetic waves that propagate without radiation along the interface between two different media. Recent investigations[2,3] on the launching of these waves have shown that the launching efficiency is a function of the reactance of the supporting surface. The excitation of radiation by these waves at a discontinuity in the reactance of the supporting surface has been studied by Brown[4] and Sharma.[5] Both the study of surface-wave launching and that of the excitation of radiation by these waves necessitates the determination of the reactance of the supporting surface. Barlow and Karbowiak[6,7] have made a detailed study of the properties of cylindrical surface waveguides. The determination of the reactance of a flat surface has been reported by Barlow and Cullen[1] and Fernando and Barlow.[8] The purpose of the present paper is to discuss the previous as well as some new theoretical and experimental approaches to the estimation of the reactance of a loss-free flat surface supporting surface waves. The experimental methods here described take into account the difficulties arising from contamination of the surface wave by the radiation field. Experimental results have been obtained on a structure supporting radial cylindrical surface waves.

(2) SURFACE-WAVE FIELDS AND THE REACTANCE OF THE SUPPORTING SURFACE

The field components of a plane surface wave supported by a flat loss-free reactive surface, described in Cartesian co-ordinates and shown in Fig. 1, are

$$H_z = A \exp(-uy - j\beta_s x) \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$E_x = A\frac{ju}{\omega\epsilon_0}\exp(-uy - j\beta_s x) \quad . \quad . \quad . \quad . \quad (2)$$

$$E_y = A\frac{\beta_s}{\beta_0}\sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)}\exp(-uy - j\beta_s x) \quad . \quad . \quad (3)$$

where
$$\beta_s^2 = \beta_0^2 + u^2 = \omega^2\mu_0\epsilon_0 + u^2 \quad . \quad . \quad . \quad . \quad (4)$$

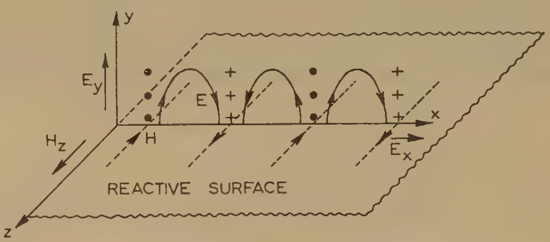


Fig. 1.—Plane surface wave.

X_s, the reactance of the surface supporting the surface-wave fields, is defined as

$$jX_s = \left(\frac{E_x}{H_z}\right)_{y=0}$$

Hence from eqns. (1) and (2),

$$X_s = \frac{u}{\omega\epsilon_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

A flat reactive surface also supports the radial cylindrical surface wave having the same physical properties as a plane surface wave, except that at large radial distances the amplitude of the radial cylindrical surface wave varies inversely as the square root of the radial distance.

(3) THEORETICAL DETERMINATION OF SURFACE REACTANCE

In practice, the reactance of a smooth metallic surface can be enhanced either by corrugating the metal or by coating the metallic surface with a layer of loss-free dielectric. Highly reactive surfaces supporting surface waves have been produced by Sharma[5] by combining these two methods.

It is intended to discuss here how the reactance of the possible types of surfaces used in practice can be calculated.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Sharma, who was formerly in the Department of Electrical Engineering, University College, London, is now in the Department of Physics, Bihar University, L.S. College, Muzaffarpur, India.

(3.1) Reactance of a Corrugated Surface

Fernando and Barlow[8] have used the expression

$$X_s = \frac{d}{D}\sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)} \tan(\beta_0 a) \quad . \quad . \quad . \quad . \quad (6)$$

to design a flat corrugated metallic surface. It is assumed that the corrugations provide a continuous reactive loading to the surface. This is justifiable only for a large number of corrugations per wavelength, a condition which can be fulfilled in practice, so that the above expression is quite useful in designing corrugated surfaces of various reactances.

(3.2) Reactance of a Coated Surface

The theoretical expression for the reactance of the interface between free space and a loss-free solid dielectric coating over a flat and smooth metallic surface, as given by Barlow and Cullen,[1] is

$$X_s = \beta_0\sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)}\frac{\epsilon_r - 1}{\epsilon_r}h \quad . \quad . \quad . \quad . \quad (7)$$

It has been assumed that h is very small and the skin depth of the metal negligible; hence this expression is useful for thin dielectric coatings only. However, it is possible to calculate the value of X_s by a simple transmission-line analogue technique without assuming that h is very small. Like the synthesis of waveguide modes in closed metallic guides from two free-space plane waves of the same frequency moving diagonally forward, it is thought that plane waves inside the solid dielectric incident on the interface between this and free space at an angle greater than the critical angle undergo successive reflections between the interface and the smooth metallic surface [Fig. 2(a)], without

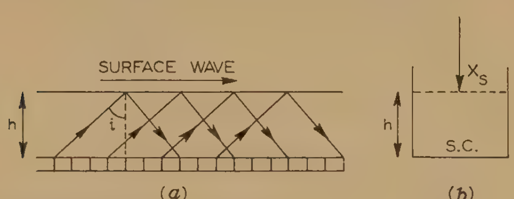


Fig. 2

(a) Paths of uniform plane-wave components in the dielectric slabs.
(b) Transmission-line analogue of (a) looking downwards.

any transmission of energy into free space. The fields above the interface are entirely reactive, decaying exponentially from the interface in the transverse direction. Looking downwards into the interface, such a system can be represented by a simple transmission line terminated at the bottom by a short-circuit and at the top by an impedance equal to the surface reactance of the interface [Fig. 2(b)]. The two terminations to the transmission line satisfy the condition of total reflection: its phase-change coefficient and characteristic impedance should be $\beta' = \beta \cos i$, and $Z_0 = Z \cos i$, respectively, where i is the angle of incidence of the plane at the interface. Hence, from transmission-line theory,

$$X_s = (Z \cos i) \tan(\beta h \cos i) \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Another expression for X_s involving the angle of incidence i can be obtained in the following manner:

The components of the propagation vector must be the same on both sides of the interface, so that

$$\beta_s = \beta \sin i \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Now, from eqns. (4) and (5),

$$X_s = \frac{u}{\omega\epsilon_0} = \frac{(\beta_s^2 - \beta_0^2)^{1/2}}{\omega\epsilon_0} \quad . \quad . \quad . \quad . \quad \text{(1}$$

Combining eqns. (9) and (10),

$$X_s = \sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)}[\epsilon_r \sin^2 i - 1]^{1/2} \quad . \quad . \quad . \quad \text{(1}$$

Hence from eqns. (8) and (11), we have

$$\tan(\beta h \cos i) = \sqrt{\epsilon_r}\frac{(\epsilon_r \sin^2 i - 1)^{1/2}}{\cos i} \quad . \quad . \quad \text{(1}$$

Thus, for given values of h, ϵ_r, and β_0, i can be graphical evaluated, and then X_s can be calculated from eqn. (11 Curve (a) in Fig. 3 has been drawn from the data calculated t

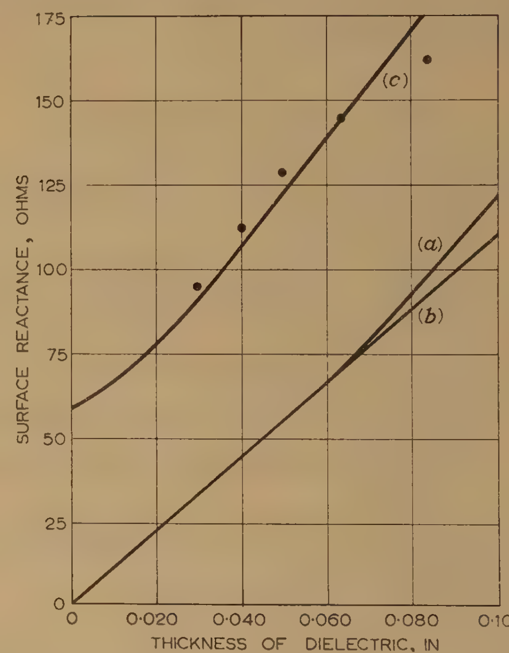


Fig. 3.—Dependence of X_s on the thickness of dielectric coating ov the surface.

(a) For coated surface: drawn from the data calculated by transmission-li analogue technique.
(b) For coated surface: drawn from data obtained from eqn. (7).
(c) For a surface obtained by coating a corrugated surface: drawn from the da calculated by transmission-line analogue technique.
● ● ● ● Experimental values of X_s for surfaces obtained by coating a corrugat surface of reactance 58 ohms with layers of Distrene of differe thicknesses.
$\epsilon_r = 2{\cdot}56$, $\beta_0 = 1{\cdot}97$ rad/cm.

this method, whereas curve (b) in the same Figure has bee drawn from the data obtained from eqn. (6). The curv demonstrate that the two methods of calculating X_s yield tl same results for small thicknesses of dielectric coating.

(3.3) Reactance of a Surface obtained by coating a Corrugat Surface

If a dielectric-coated structure is used, the reactance of t surface can be enhanced by increasing the thickness of t coating, and if a corrugated structure is used, this can be do by increasing the corrugation depth. An alternative method to put a dielectric coating over the corrugated surface. Such device has been used by Sharma[5] to study the excitation radiation by surface waves at a discontinuity in surface rea

ance. The surface reactance at the interface between free space and the dielectric coating can be calculated in this case by the same transmission-line analogue technique as described in Section 3.2. It is necessary to assume that the corrugations provide a uniformly distributed reactance at the metallic surface, and that the relevant transmission-line analogue has a terminating impedance at the bottom which is equal to the reactance of the corrugated surface. The transmission-line analogue for

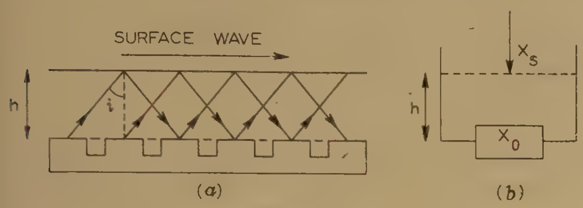


Fig. 4

(*a*) Paths of uniform plane-wave components in the dielectric slab.
(*b*) Transmission-line analogue of (*a*) looking downwards.

this case will be as shown in Fig. 4(*b*), where X_0 is the reactance of the corrugated surface. The two terminations to the transmission line satisfy the condition of total reflection in this case also.

On transmission-line theory X_s will be given by

$$X_s = (Z \cos i)\frac{X_0 + (Z \cos i) \tan (\beta h \cos i)}{(Z \cos i) - X_0 \tan (\beta h \cos i)} \quad . \quad (13)$$

combining this equation with the equation

$$X_s = \sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)}[\epsilon_r \sin^2 i - 1)^{1/2}$$

which follows in the same way as discussed in Section (3.2), we have

$$\tan (\beta h \cos i) = \frac{\cos i[\sqrt{\epsilon_r}(\epsilon_r \sin^2 i - 1)^{1/2} - X][\cos^2 i - X\sqrt{\epsilon_r}(\epsilon_r \sin^2 i - 1)^{1/2}]}{\cos^4 i - X^2\epsilon_r(\epsilon_r \sin^2 i - 1)} \quad . \quad . \quad . \quad . \quad (14)$$

where $X = X_0/Z$.

The angle of incidence i can be graphically solved from this equation, so that X_s can be evaluated from eqn. (11).

The results of numerical evaluation of X_s are shown in curve (*c*) of Fig. 3.

(4) EXPERIMENTAL ESTIMATION OF SURFACE REACTANCE

(4.1) Theory

It has been mentioned in Section 2.1 that

$$X_s = \frac{u}{\omega\epsilon_0}$$

Further, the relation between X_s and β_s, and that between X_s and λ_s as obtained from eqns. (4) and (5) are given by

$$X_s = \sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)}\left[\left(\frac{\beta_s}{\beta_0}\right)^2 - 1\right]^{1/2} \quad . \quad . \quad . \quad (15)$$

and

$$X_s = \sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)}\left[\left(\frac{\lambda}{\lambda_s}\right)^2 - 1\right]^{1/2} \quad . \quad . \quad . \quad . \quad (16)$$

Thus an experimental estimation of any of the three quantities, u, λ_s, and β_s will determine the value of X_s. A discussion of the methods for determining u, λ_s, and β_s will therefore be the subject of this Section.

(4.1.1) Determination of *u*.

The vertical decay coefficient, u, can be determined by an experimental arrangement, described by Fernando and Barlow,[8] which measures vertical field distribution above the reactive surface. This gives an accurate value of u provided that the radiation field contamination in the surface wave is not large, and the vertical decay of the field is not too rapid to be measured accurately. Hence the estimation of u helps to determine X_s under limited conditions only.

(4.1.2) Determination of λ_s and β_s.

The guide wavelength, λ_s, can be evaluated by terminating the guiding structure by a short-circuiting plate. For a coated surface the standing-wave pattern can be accurately obtained by means of a standing-wavemeter arrangement. However, for structures using corrugated metallic sheets, because of the distortion of the field near the corrugations, it will be necessary to measure the fields only at the centres of the ridges and corrugations. Thus only a few discrete points will be available per wavelength. In spite of this difficulty, the standing-wave pattern can be mapped approximately, and λ_s can be determined by measuring the distance between, say, ten to eleven minima of the pattern. This method is similar to that discussed by Mullett and Loach[9] for measuring the guide wavelength of a circular corrugated guide. Fig. 5 shows a typical standing-wave pattern

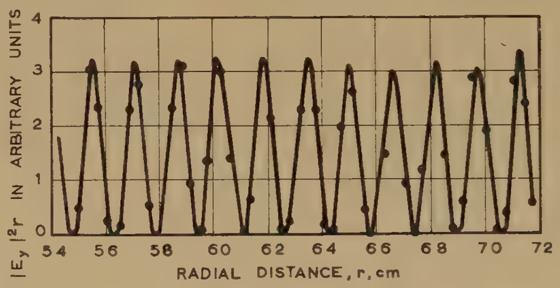


Fig. 5.—Standing-wave pattern over the corrugated surface, of reactance 58 ohms.

for a corrugated surface obtained at a great distance from the launcher. The radial cylindrical type of surface wave has been used to obtain the standing-wave pattern. The pattern shows that the maxima are fluctuating. However, the minima are hardly distinguishable from zero along the pattern shown in Fig. 5. It is reasonable to assume that the experimentally determined pattern is formed by the combination of the standing wave produced by the guided surface wave and the wave radiated from the launching aerial. Further, if it is also assumed that the field strength of the radiated wave is very small in the vicinity of the surface compared with the surface-wave field, then it can be shown that the maxima and minima of the pattern may be represented by the equations:

$$\left.\begin{aligned} r_{max}|E|^2_{max} &\simeq |A|^2\left(1 + \frac{b}{\gamma^{1/2}_{max}} \cos \beta_0\gamma_{max}\right) \\ r_{min}|E|^2_{min} &\simeq 0 \end{aligned}\right\} \quad . \quad (17)$$

where $B/A = b$, B being the amplitude factor of the radiated wave, and r_{max} and r_{min} radial distances of the standing wave maxima and minima respectively.

These expressions indicate sufficiently accurately the behaviour of the maxima and minima of the experimentally obtained standing-wave pattern. It is clear that the minima remain unaffected by the small radiation field contamination at the surface, so that λ_s can be obtained from the standing-wave pattern without loss of accuracy. The values of X_s determined

from the value of λ_s for the case of a guiding structure obtained by coating a corrugated surface with dielectric sheets (Distrene) of different thicknesses are shown in Fig. 3. The theoretical and experimental results agree for small thicknesses of the coating; their discrepancy for thicker coatings indicates that the assumptions made for theoretical calculations are justifiable only for thinner coatings. The author has used this method of estimating surface reactance in his investigation of the excitation of radiation by surface waves at a discontinuity in surface reactance. At great distances from the discontinuity it has been found that the wave radiated from the discontinuity does not produce any measurable effect on the minima of the surface-wave standing-wave pattern measured at the surface.

Instead of using a short-circuited guiding structure and determining λ_s, a structure terminated by a matched load can be used, and β_s can be determined by measuring the phase distribution at the guiding surface in the direction of propagation. A phase bridge of the type shown in Fig. 6 was used for this purpose. A

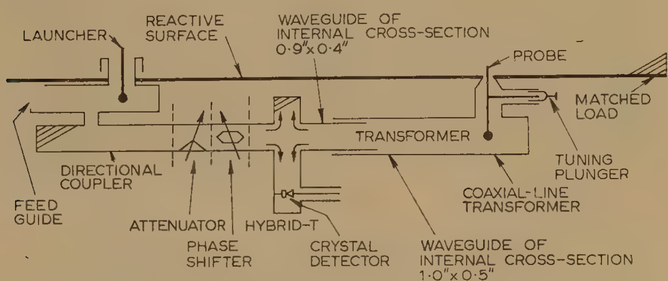


Fig. 6.—Phase bridge to study the phase distribution in the direction of propagation of surface waves.

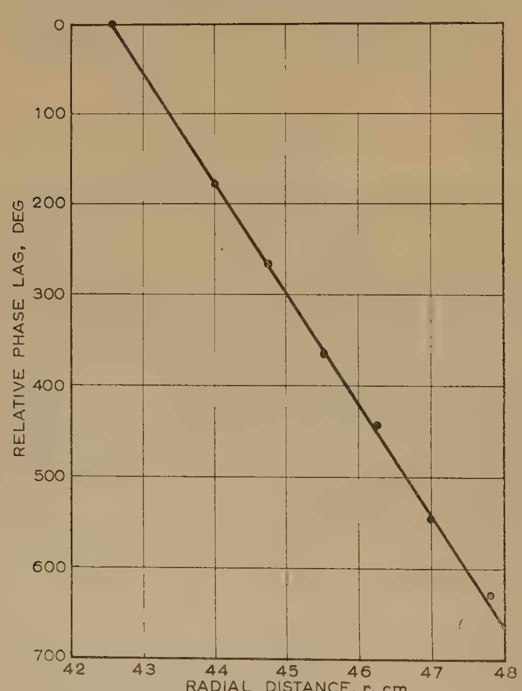


Fig. 7.—Phase distribution along the direction of propagation.

typical phase distribution curve is shown in Fig. 7. The val of β_s measured by this method, for a guiding structure in whi a corrugated surface of reactance 58 ohms was coated with 0·050 in Distrene sheet, has been found to be 119·7°/cm. T value of X_s calculated from this value of β_s is 129·0 ohm which agrees with the value obtained by determining λ_s.

(5) CONCLUSIONS

The experiments described demonstrate that the reactance loss-free flat surfaces supporting surface waves can be estimat despite the difficulties usually met in practice. Within the limi of experimental error and the validity of the assumptions ma for theoretical calculations, the agreement between theoretic calculations and experimental results justifies the application the transmission-line analogue technique to theoretical calcul tions. It is hoped that the methods of estimating the reactan of a loss-free flat surface here discussed will be helpful in studyi the various problems in the field of surface waves.

(6) ACKNOWLEDGMENTS

The experimental work was carried out in the Electric Engineering Department of University College, London. T author wishes to express his gratitude to Prof. H. E. N. Barlo for the laboratory facilities provided, and to Dr. J. Brown f the benefit of many stimulating discussions. He also wishes thank Dr. S. P. Sinha, Professor of Physics, L.S. Colleg Muzaffarpur, for his encouragement during the preparation the paper.

(7) REFERENCES

(1) BARLOW, H. M., and CULLEN, A. L.: 'Surface Wave *Proceedings I.E.E.*, Paper No. 1482, April, 1953 (**10** Part III, p. 329).
(2) CULLEN, A. L.: 'The Excitation of Plane Surface Wave *ibid.*, Monograph No. 93, February, 1954 (**101**, Part I p. 1).
(3) CULLEN, A. L.: 'A Note on the Excitation of Surface Wave *ibid.*, Monograph No. 239 R, May, 1957 (**104 C**, p. 47
(4) BROWN, J.: 'The Behaviour of Surface Waves at a Disco tinuity in Surface Reactance' (to be published).
(5) SHARMA, K. P.: 'An Investigation of the Excitation Radiation by Surface Waves', *Proceedings I.E.E.*, Pap No. 2737 E, March, 1959 (**106 B**, p. 116).
(6) BARLOW, H. M., and KARBOWIAK, A. E.: 'An Investigatio of the Characteristic of Cylindrical Surface Waves', *ibi* Paper No. 1462, November, 1953 (**100**, Part III, p. 321).
(7) BARLOW, H. M., and KARBOWIAK, A. E.: 'An Experiment Investigation of the Properties of Corrugated Cylindric Surface Waveguides', *ibid.*, Paper No. 1625, May, 19 (**101**, Part III, p. 182).
(8) FERNANDO, W. M. G., and BARLOW, H. M.: 'An Investig tion of the Properties of Radial Cylindrical Surface Wav Launched over Flat Reactive Surfaces', *ibid.*, Pap No. 2009, May, 1956 (**103 B**, p. 307).
(9) MULLETT, L. B., and LOACH, B. G.: 'Experimental Work Corrugated Waveguides and Associated Components f Linear Electron Accelerators', *Proceedings of the Physic Society*, B, **61**, p. 271.

621.372.8: 621.744

The Institution of Electrical Engineers
Paper No. 2996 E
July 1959
©

MANUFACTURE OF WAVEGUIDE PARTS BY INVESTMENT CASTING FROM FROZEN-MERCURY PATTERNS

By H. H. SCHOLEFIELD, Ph.D., B.Met., F.Inst.P., H. H. H. GREEN, B.Sc.(Eng.), Associate Member, and R. E. GOSSETT.

(*The paper was first received* 21*st January, and in revised form* 24*th March*, 1959.)

SUMMARY

The paper explains and discusses the capabilities and limitations of the recently developed frozen-mercury process for the production of cast parts.

Many types of waveguide, including bends, horns, mixers, etc., are currently being manufactured, and some of these incorporate structural support features. As flanges are cast on, the only finishing necessary is the machining of the flange faces, including any choke grooves and the drilling of fixing holes. Cast complex horns containing transfer sections have proved particularly valuable as they are very difficult to produce by fabrication techniques. The waveguide parts are cast in aluminium alloy, gunmetal or other alloys, according to requirements. Consistency of dimensions in relationship to design tolerances are examined. Typical examples are described in the paper and tolerances and surface finish obtainable discussed.

(1) INTRODUCTION

The frozen-mercury process was developed about ten years ago in the United States and has been extensively employed in the manufacture of waveguide components both there and in this country.* The use of the frozen-mercury process has overcome many problems and difficulties associated with the production of waveguide parts, particularly in maintaining dimensional accuracy and required surface finish. It also allows parts incorporating internal irises and structural features to be produced and ensures sound castings. Waveguide parts should be designed to take advantage of the facilities which the process offers. Parallel-sided cavities can be produced without draft, and this and the ability to produce aluminium-alloy castings having a wall only 0·08 in thick are of particular value.

The fact that patterns are free from shrinkage is an improvement over other techniques which employ an expendable pattern material.

(2) DESCRIPTION OF PROCESS

The process is an investment-casting technique, using mercury as the expendable pattern material. In investment casting a positive expendable pattern (e.g. of wax or mercury) of the object required is surrounded with a refractory slurry. The slurry is allowed to set and the pattern removed by melting and pouring it out, or by dissolving it, so leaving a cavity of the required shape in refractory. This is then fired and the metal poured into the cavity, so producing the actual casting required.

The use of mercury as against other materials confers a number of advantages:

(*a*) Solid/liquid metal expansion is small, permitting shell moulds to be employed.

(*b*) It readily welds to itself under light pressure, so that complex patterns can be produced by abutting simple component patterns together on jigs, which ensure accurate alignment.

(*c*) The high-density and good wetting characteristics of the pattern material permit accurate reproduction of minute detail to be achieved.

A steel or aluminium die is employed for producing the pattern, being designed in such a way that the frozen-mercury pattern can be withdrawn from it without distortion. Pre-hardened steel is used to minimize damage due to handling during pattern manufacture. The die cavities undergo practically no wear and consequently do not require renewal, although occasionally pins and bushes need to be replaced. In the case of aluminium dies their lightness reduces the possibility of damage during handling and wear is minimized by hard anodizing.

Since dimensional accuracy of the mercury pattern is directly related to that of the die it is obvious that the die must be machined to close tolerances and the component parts must also be accurately fitted to prevent mercury leakage.

The die is filled with mercury and immersed in a refrigerated bath at about $-65°$ C. The mercury freezes and the pattern is extracted from the die. The pattern is a facsimile of the finished casting required, with the requisite dimensional allowances for the changes in dimensions of the die during cooling down to the freezing point of mercury and the changes which occur during solidification and cooling of the casting. The pattern is invested with a zircon-base ceramic shell mould to a thickness of up to $\frac{1}{4}$ in. The mercury is allowed to melt out at room temperature, leaving a 'green' shell, the interior of which accurately reproduces every detail of the pattern. The 'green' mould is fired at high temperatures to render it rigid, stable and inert to most liquid metals. In this form, the mould is permeable and strong enough to withstand various casting methods, including, for example, centrifugal, vibration, suction and sling casting, which are employed in the production of metallurgically sound and dimensionally accurate castings.

A great advantage of the frozen-mercury process is that a shell mould is employed. This permits chills and other methods to be used, allowing directional solidification to be induced. As a consequence, it is possible to obtain sound castings of the required metallurgical structure. Aluminium, copper, magnesium and iron alloys can be cast into the moulds without any significant mould-metal reaction occurring.

(3) TOLERANCES AND SURFACE FINISH

Microwave components call for two approaches to the dimensional tolerance problem, one from the mechanical standpoint and one from the electrical-performance requirements. To a large extent the tolerances from the latter dictate the methods of manufacture that are possible. The commonly occurring tolerance of $\pm 0 \cdot 001$ in, which represents the absolute limit of dimensional accuracy which the authors have been able to achieve in cast components, has been a factor militating against their use.

* SCHOLEFIELD, H. H.: 'Investment Casting, using the Mercury Pattern Method', Institution of Mechanical Engineers Conference on Technology of Engineering Manufacture, March, 1958.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Scholefield and Mr. Gossett are with Sankey-Telcon Ltd., and Mr. Green is with Standard Telephones and Cables Ltd.

It should be emphasized that to obtain a satisfactory component using the frozen-mercury process the design should be drawn up with casting specifically in mind. It is unwise to base designs on drawn tube and sheet material and then expect casting methods to reproduce the same outlines. In many cases absolute dimensions are not of such great importance as consistency. A designer may ask for a dimension, e.g. 1·548 ± 0·001 in, when, in fact, if a casting can be made to give a dimension within the range 1·548 ± 0·005 in he will be satisfied, provided that whatever dimension emerges is consistent from casting to casting with a tolerance of ±0·001 in. These points are not always realized and are really a basis for discussion between designer and manufacturer. After having gained experience with cast waveguide components, engineers can design accordingly, and consequently, if repetition can be guaranteed, then very often successive modifications of the dies to achieve close absolute dimensions are unnecessary.

The surface finish attained on aluminium-alloy waveguides is $3{\cdot}4 \times 10^{-5}$ in (r.m.s.) after vapour blasting. The 'as cast' surface is correspondingly better. For copper alloys a surface of $5\text{–}6 \times 10^{-5}$ in (r.m.s.) in the vapour blasted condition is attainable. It has been found* that the surface roughness may be related to the attenuation by:

Multiplying factor ..	1	1·2	1·6	1·8
Surface roughness ..	0	$\frac{1}{2}\delta$	δ	2δ

where δ is the skin depth. The skin depth at 10 Gc/s in copper is 29×10^{-6} in and in aluminium 49×10^{-6} in. It would therefore appear that the absolute dimensional tolerances and surface roughness on aluminium-alloy frozen-mercury components are such that the limit of application would be at 10 Gc/s or less. There are, however, other factors to be considered which make frozen-mercury castings well worth while at 10 Gc/s and above. Firstly, under certain circumstances it is possible to relax tolerances appreciably, and secondly, particularly in production quantities, it is difficult to find a more satisfactory manufacturing method. It is not possible to generalize, and the authors consider that each case should be treated on its merits. It is not proposed to go into tolerancing problems generally, since this is outside the scope of the paper.

In the manufacture of microwave components from drawn tube or sheet material the surface roughness of the tube as usually supplied is less than 2×10^{-5} in, and the inside dimensions for the 10 Gc/s band waveguide are within ±0·001 in, on nominal. It is unlikely that fabrication of components of any complexity will leave the surface roughness unchanged or leave the cross-section undistorted. This is particularly true at brazed joints and small-radius bends. Unless great care is taken, the variance associated with fabricated components, from both mechanical and electrical requirements, tends to be large. The conclusion, reached on a limited number of components, results for which are quoted later, is that there is little difference in performance between a fabricated assembly having a high effective surface roughness and a frozen-mercury casting having a high intrinsic surface roughness. The latter has, moreover, the asset of smaller variance.

Pursuing further the possibility of tolerance relaxation from a theoretical aspect, we may write

$$Z_w = \frac{a}{b}\sqrt{\frac{\mu}{\epsilon}}\frac{1}{\sqrt{\left[1 - \left(\frac{\lambda_0}{2b}\right)^2\right]}} \quad . \quad . \quad . \quad . \quad (1)$$

where Z_w is the wave impedance as usually defined for the H_{01} mode in rectangular waveguide and the other symbols have th usual significance. In tolerancing, variations in both a and must be taken into account, i.e.

$$\Delta Z_w = \frac{\partial Z_w}{\partial a}da + \frac{\partial Z_w}{\partial b}db \quad . \quad . \quad . \quad . \quad .$$

which, for small variations, may be written

$$\Delta Z_w = \frac{Z_w}{a}\Delta a - \frac{Z_w \Delta b}{b\left[1 - \left(\frac{\lambda_0}{2b}\right)^2\right]} \quad . \quad . \quad .$$

A criterion for variations in Z_w can be specified only particular cases. For good performance in the 10 Gc/s ba for example, a v.s.w.r. not greater than 1·02 is sometim specified. The use of eqn. (3) should be confined to sm deviations not only to conform with its derivation but a because for large deviations discontinuity effects could beco an appreciable factor.

Taking the worst limiting case for the H_{01} mode in No. waveguide in which the variations in a and b are equal a opposite:

$$\frac{\Delta Z_w}{Z_w} = \frac{\Delta a}{a} + \frac{\Delta b}{b\left[1 - \left(\frac{\lambda_0}{2b}\right)^2\right]} \quad . \quad . \quad . \quad .$$

Inserting the values $a = 0{\cdot}400$ in, $b = 0{\cdot}900$ in, $(\lambda_0/2b)^2 =$ (and $\Delta Z_w/Z_x = 0{\cdot}02$,

we have $$\Delta a = \Delta b = 0{\cdot}004\,2\text{ in} \quad . \quad . \quad . \quad . \quad .$$

—a rather larger tolerance than is customary.

It may be agreed that this represents only one type of dime sional variation, namely the uniform change of a cross-secti dimension. This is, of course, true, but in the authors' experien this is the most commonly occurring dimensional shift in t frozen-mercury process. To facilitate production the toleran permitted on a waveguide mixer unit casting was increased fro ±0·004 to ±0·008, when it was proved that the dimensio discrepancies occurred over a gradually tapering area, and n in localized areas giving rise to a wavy surface. In a particu case, a number of types of bend were designed for a mean pa length, an integral number of half guide-wavelengths at 8 900 M in No. 16 waveguide cross-section. The measured v.s.w.r. grade 1 equipment on sixty bends gave values lying between 1· and 1·02. Performance figures of this consistency had o previously been achieved by precision machined compone milled and turned from the solid. Components produced tube bending and brazing gave v.s.w.r. figures distributed o the range 1·02–1·10.

A slightly different problem on bend design arose where E-plane 90° bend was required with a very small radius to m mechanical requirements. The mismatch was compensated an inductive post. Figures for 20 bends gave a spread of v.s.w values from 1·00 to 1·03. In this instance it proved impossi to fabricate consistently to this standard.

In all cases above, it was noted that the measurable gu cross-sections were consistent to within ±0·001 although t was not specifically required by the drawings. The conclusi we have drawn is that, in many cases, it is possible to re tolerances considerably without loss of performance.

(4) SOME TYPICAL CASTINGS

The types of waveguide components which have been produ by this process, shown in Figs. 1 and 2, include E- and H-pl

* Harvey, A. F.: 'A Surface-Texture Comparator for Microwave Structure', *Proceedings I.E.E.*, Paper No. 1756 R, March, 1955 (102 B, p. 219).

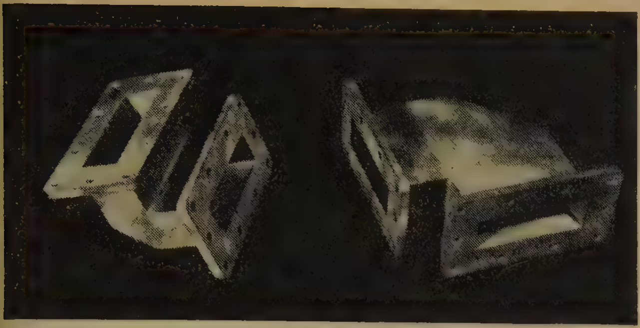


Fig. 1.—E- and H-plane bends.

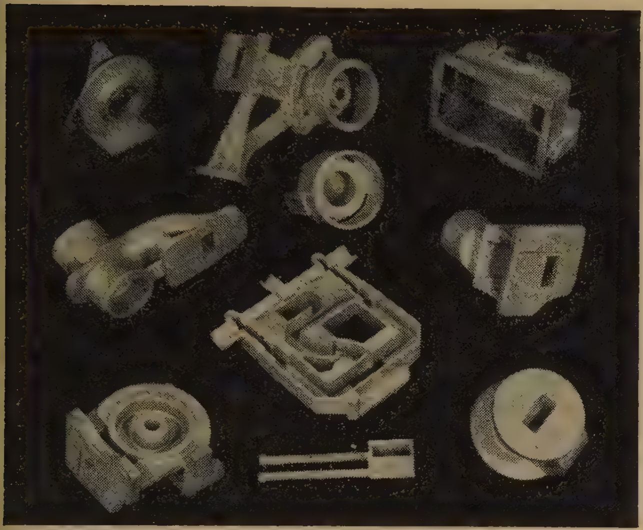


Fig. 2.—Waveguides incorporating irises, mixer units, rotating joints, duplexers and crystal holders in aluminium and copper alloys.

bends, waveguides incorporating irises, mixer units, rotating joints, duplexers, crystal holders, klystron mounts and horn feeds in both 3 and 10 Gc/s bands, in aluminium and copper alloys.

An H-plane-bend 10 Gc/s-band waveguide component is shown in Fig. 3. This bend was actually cast in aluminium alloy LM8, but could have been produced in other aluminium alloys, in gunmetal, beryllium-copper or many other non-ferrous alloys including magnesium, provided that the relevant pattern-maker's contraction had been allowed for in the die. Should the need arise for waveguide components in guided missiles, say of stainless steel or heat-resisting steel, the parts could also be cast in these alloys. Dimensional tolerances of $\pm 0 \cdot 002$ in in the waveguide aperture with zero draft and a wall thickness of $0 \cdot 080$ in were maintained. The only machining operation required was on the external flange faces.

Some of the advantages of producing this type of part by this technique are:

(*a*) The flanges are cast integrally with the bend. In the case of flanges soldered or brazed to fabricated bends, the design, of necessity, incorporates thickened sections at the flanges, whereas a simple fillet radius is all that is necessary in the case of cast components. The inclusion of the flanges also aids the casting technique since they can be used as gating faces and give rigidity to the pattern.

(*b*) As shown in Fig. 4, close angle bends which cannot easily be fabricated can be produced.

(*c*) The fact that soldering and brazing are not used means that no distortion arises owing to these operations.

(*d*) Castings have greater rigidity, which is of value where the components are also structural parts.

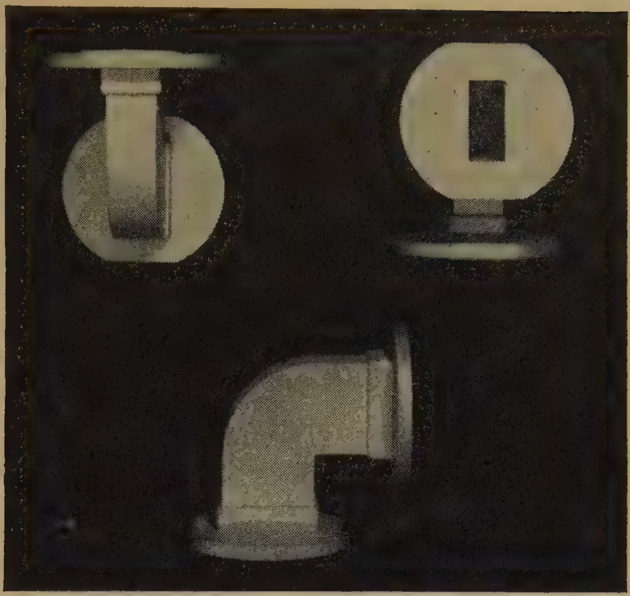


Fig. 3.—H-plane 10 Gc/s bend.

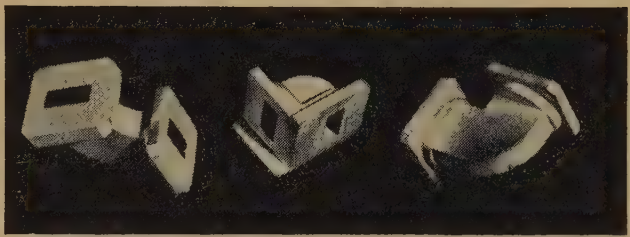


Fig. 4.—Close angle bends.

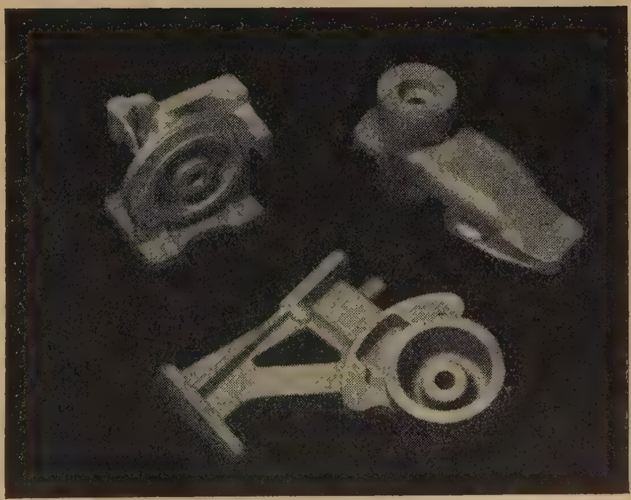


Fig. 5.—Waveguides designed to achieve maximum strength at minimum weight.

Fig. 5 shows examples of waveguides designed to achieve maximum strength at minimum weight, and also one in which the fixing bracket is cast integrally with the guide.

The facility for incorporating such features is of considerable

advantage, conferring saving in space and economy in manufacture. In fabrication, it is necessary to incorporate flanges, bosses, etc., which can be omitted when the component is produced as an integral casting.

(5) CONCLUSIONS

It will be realized that the process has substantial advantages to offer over other methods of producing waveguide components. It should be emphasized that the frozen-mercury process is primarily intended as a production process in this context for microwave components working at 10 Gc/s or lower frequency and of normal commercial performance. It is not intended for manufacture of ultra-high precision ($\pm 0{\cdot}0005$ in), 4×10^{-6} in (r.m.s.) surface-finish components required for research, te equipment and millimetric work.

Die costs are undoubtedly high but by no means prohibitiv and, indeed, when reasonable runs are required die costs becom of secondary importance. The component cost is markedly les than that for conventional machined parts and is more tha competitive with any other method of production which ca achieve comparable results. The dimensional consistency an high surface finish and the ability to incorporate structur features result in substantial economies as well as offering greate freedom to the designer.

Encouraging co-operation has been received both fro industry and many Government establishments.

DISCUSSION ON
'A NOVEL, HIGH-ACCURACY CIRCUIT FOR THE MEASUREMENT OF IMPEDANCE IN THE A.F., R.F. AND V.H.F. RANGES'*

Communications on this paper have been received from **Messrs. M. V. Callendar; L. G. Cripps and M. J. Gay; T. A. Deacon, L. H. Ford and J. J. Hill; A. C. Lynch; M. Myers; G. H. Rayner; A. H. Silcocks; J. K. Webb; and D. Woods.**

In these contributions there is a substantial amount of common ground, the main points of which may be summarized as follows:

(1) The claim that the circuit is novel is unjustified, since it is in essence a conventional form of voltage-dividing network. It is, in fact, merely a rearrangement of the so-called transformer bridge or a.c. potentiometer. All that is new is the disposition of the elements in the secondary circuits of the transformer.

(2) There is an error of sign on the right-hand side of eqn. (1), the consequences of which run right through the mathematics and the interpretation in terms of balance conditions. It implies, in fact, that the bridge could not possibly balance, for it is obvious from simple physical principles that, for a balance, M_1 must be numerically greater than M and of opposite sign instead of the condition given in the paper.

One can only assume that, in writing up the mathematics for publication, some carelessness occurred or that the diagrams have been labelled differently from those used in deriving the mathematics, and that the equations actually used in deducing the results from the experimental measurements were the correct ones and not those so misleadingly given in the paper.

(3) The discussion about the neglect of the various capacitances to earth is in principle fallacious. For instance to state that C_2 and C_3 do not matter because Z_3 does not appear in the balance equations ignores the fact that if C_2 and C_3 are taken into account there is current flowing in Z_2 and in C_2 and C_3 at the balance condition when no current flows through the detector. This follows quite simply from network theory and it results in the fact that the balance conditions contain a term including Z_3, C_2 and C_3 which disappears only when C_2 and C_3 are zero. Similar criticisms are made of the statement that the capacitances C_1 and C_4 shunt the impedance Z_1, where a full analysis of the circuit shows that the balance condition is affected by the values of C_1 and C_4.

(4) The staff of the National Physical Laboratory state that the accuracy of certification for capacitors and inductors is not normally better than 0·01%, and they are authorized to state that at no time has the value of a capacitor been certified by the N.P.L. with an accuracy of 0·001%, and that no capacitor of which the N.P.L. has had experience has shown a constancy of this precision over an appreciable period of time.

(5) Many comments are made on various factors which are known by experts in this field to need very careful attention if accuracies approaching those quoted in the paper are to be achieved, but which are ignored or dismissed by saying that their effects can easily be taken into account. It is most unfortunate that the final Section on experimental verification and the Appendix are so brief. If the author indeed took special precautions in his experimental set-up and in making sure that the stray effects, etc., which he ignores really are negligible, it would have been most valuable to have a careful discussion on this aspect to justify the sweeping claims he makes for his circuit.

As it is, the reader is faced with a mathematical argument that has to be completely rewritten if it is to apply to the actual bridge as used, together with statements with regard to capacitances to earth that are in principle wrong and in spite of which the bridge gives results some of which are accurate to limits which are beyond those which can actually be certified by the N.P.L.

If, by his arrangement of the bridge and the experimental techniques he used, the author actually achieved a method that is really an advance on existing procedures, it would be most useful to have a careful description that will convince the experts who are familiar with the difficulties of making such accurate measurements at the very high frequencies claimed.

Dr. D. Karo (*in reply*):

(1) The circuit is novel. The comparison with an a.c. potentiometer is surely not serious—it would in fact be the same as saying that a modern d.c. machine is the same as Barlow's disc.

To say that the circuit is a voltage-dividing network is quite superfluous; every impedance-measuring circuit since Ohm's law was first written is in a sense a voltage-dividing network.

As for the comparison with a transformer bridge, apart from the new disposition of the circuit-elements in the secondary circuits of the transformers, most of the balance equations are independent of frequency, there is no output transformer, and all the earth capacitances are eliminated in a practical way, which is not the case in the transformer bridge.

(2) A regrettable error resulted from the misplacement of a minus sign in the first of eqns. (1); this minus sign ought to be on the left of the equation, in front of $jM_1\omega I_p$ and not on the right-hand side, in front of Z_2I_1. This error is immaterial to the calculations, the magnitudes of the test results being unaffected by it. One can only assume that the contributors to the discussion have not checked the mathematics, since they would surely have discovered that the error is of no importance as far as the practical results are concerned. A list of corrections will, however, be found at the end of this reply.

(3) The author respectfully suggests that the contributors re-read the paper and try out the circuit with suitable components, since they raise objections to errors and difficulties that are non-existent. The manner in which the capacitances to earth are dealt with is entirely justified. The effect of any additional current in Z_3 (Fig. 1) due to the capacitances C_2, C_3, C_1 and C_4 (Fig. 4) can be treated either as a change in Z_3 or as an impurity of M (or N); this treatment is well known and has been exhaustively dealt with by several authorities.

In the paper it is clearly stated that C_2 and C_3 alter the impedance Z_3, whilst the effect of the other earth capacitances is treated as an impurity in M (or N).

Let the additional current in the secondary of M due to any relevant earth capacitances be I and $jM'\omega I_p$ be the e.m.f. in the secondary of M when $I = 0$; when I is not zero one can write $jM'\omega I_p - Z_3I = jM\omega I_p - Z_{3e}I \simeq jM\omega I_p$ or $jM'\omega I_p - Z_3I = jM\omega I_p \pm \sigma I_p$ according as to whether one considers a change in impedance or an impurity, M being the effective value of the mutual inductance, Z_{3e} the effective impedance and σ the phase defect.

$Z_{3e}I$ is nearly in phase with R_3I and negligibly small, since at

* Karo, D.: Paper No. 2709 R, November, 1958 (see 105 B, p. 505).

high frequencies R_3I is negligibly small and I_p is nearly in phase with I; this is evident from the most elementary series-circuit vector diagram.

In other words: since at balance the only current in Z_3 is I, we have at high frequency a current flowing through a capacitance and inductance in series; the result is, of course, a change in the impedance and an effective value of M.

The effect of σ in the circuit is clearly explained in Section 4; it is also clearly stated that N_1 and N are the effective numbers of turns in the h.f. transformers.

Obviously, since Z_3 does not otherwise appear in the balance equations, this is the simplest way of dealing with the relevant earth capacitances; especially so, since opposing M_1 and M through known standards allows the calibration of M for any frequency and any earth capacitances.

Let the impedance of C_1 and C_4 be Z and the current through it I'; the current through Z_4 is then $I + I'$, and as at balance $ZI' = Z_1I_1$, we have $I' = Z_1I_1/Z$. The drop in Z_4 is therefore $(Z_4 + Z_4Z_1/Z)I_1$ and in the first of eqns. (1) one can write,

$$-Z_4I_1 - I_1Z_4Z_1/Z - Z_1I_1 = -Z_4I_1 - Z_1I_1(1 + Z_4/Z)$$

instead of $-Z_4I_1 - Z_1I_1$.

The scale of Z_1 ought therefore to be multiplied by $(1 + Z_4/Z)$; this is the meaning of the statement in the paper that the scale of Z_1 should be calibrated to include the capacitances C_1 and C_4.

Since at high frequencies $Z_4 \simeq jL_4\omega$, the multiplying factor will be real and the calibration of the scale of Z_1 easy and practical.

Surely this explanation is too elementary to have warranted inclusion in the *Proceedings*.

Since Z_4 appears in the main balance equations, the treatment of C_1 and C_4 as an impurity in M_1 would be less practical.

(4) The standards used, as well as the oscillators, were certified by manufacturers' signed calibration tables. The author was also informed that the equipment was certified for the manufacturers by the N.P.L., and this information was accepted in good faith. If any part of the equipment was not certified, or certified to limits different from those shown on the manufacturers' certificates, the author tenders his apologies to the N.P.L.

The capacitor of Table 1 was definitely certified by the manufacturers to 0·001%; I am in possession of three such certificates; whether the calibration is stable in time is beside the point.

(5) The factors mentioned can easily be taken into accou[nt] and no expertize is necessary; every graduate student takin[g] electrical measurements is familiar with them. All that [is] required is proper screening, judicious disposition of the equip[-]ment, use of screened coaxial cables, adherence to manufacturer[s'] indications, calibration of the mutual inductors or transforme[rs] for given earth capacitances and frequency, and the calibratio[n] of Z_1 in the relevant circuits of Figs. 10 and 11.

The mathematical treatment, except for the error in sign whic[h] is of no practical importance, is correct; it need not be rewritte[n] and the treatment of the earth capacitances is entirely justifie[d].

I see no reason why results better than those achieved b[y] any laboratory whatever cannot be achieved if one uses a bette[r] and more suitable circuit.

I would welcome any queries as to the details of the techniqu[e] used.

CORRIGENDA

Pages 506 *et seq.*

Eqns. (1): a minus sign *should precede* $jM_1\omega I_p$ and the minu[s] sign before Z_2I_1 *should be deleted*, in the first o[f] these two equations.

Eqns. (2) *should read*

$$-\frac{M_1}{M} = \frac{Z_2 + Z_1 + Z_4}{-Z_2}; \quad \frac{M - M_1}{M}Z_2 = -Z_4 - Z_1$$

In the balance equations,

$N_1 - N$ and $N > N_1$ *should read* $N - N_1$ and $N_1 > N$

and

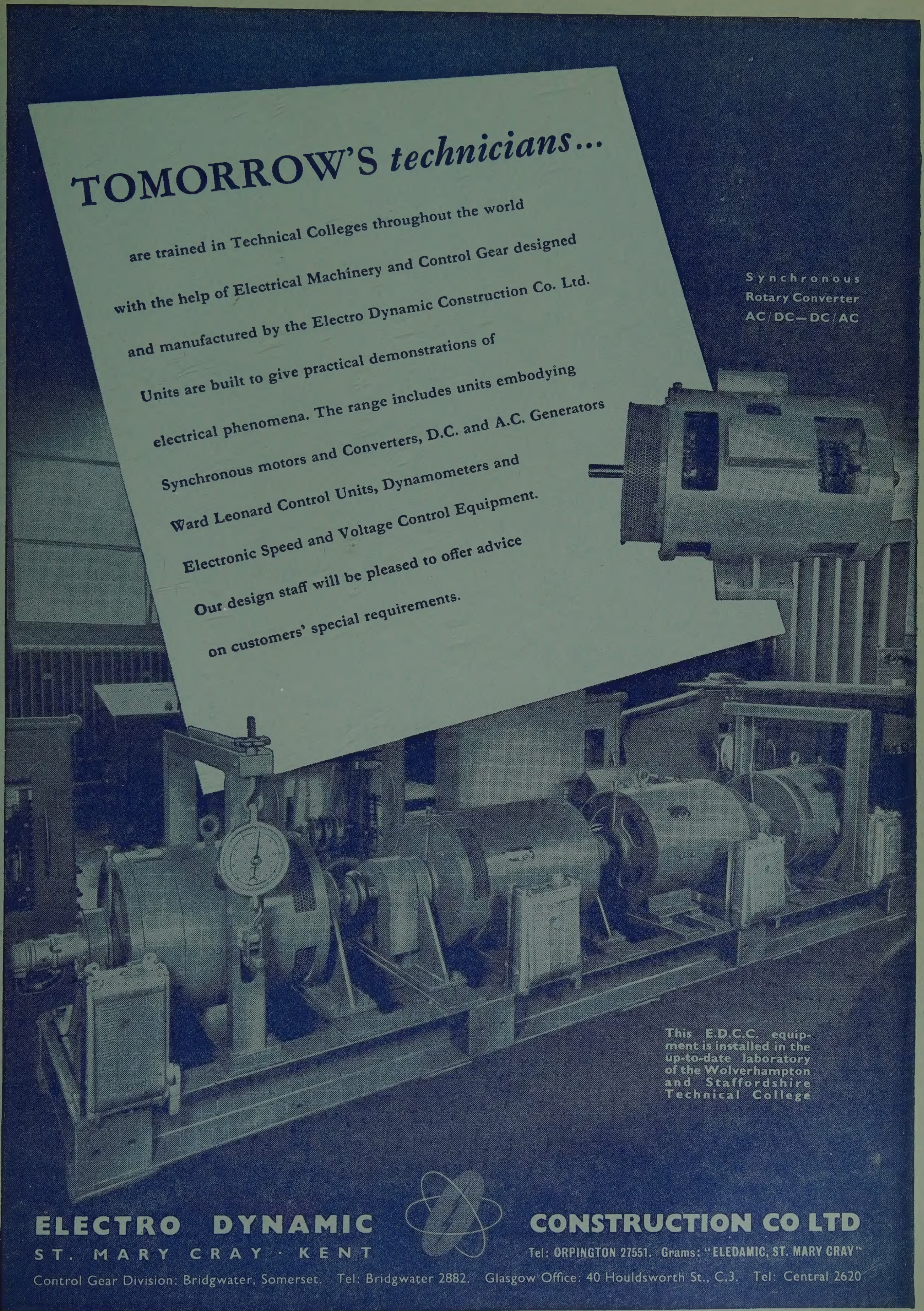
$M_1 - M$, $M - M_1$ and $M > M_1$ *should read* $M - M_1$, $M_1 - M$ and $M_1 > M$.

Page 509, Table 2, under 'Certified values',

$R_x + 2{\cdot}22\Omega \pm 0{\cdot}1\%$ *should read* $R_x = 2{\cdot}22\Omega \pm 0{\cdot}1\%$

and under 'Result of test',

$250 + 10^3\Omega$ *should read* $250 \times 10^3\Omega$.


TOMORROW'S technicians...
are trained in Technical Colleges throughout the world
with the help of Electrical Machinery and Control Gear designed
and manufactured by the Electro Dynamic Construction Co. Ltd.
Units are built to give practical demonstrations of
electrical phenomena. The range includes units embodying
Synchronous motors and Converters, D.C. and A.C. Generators
Ward Leonard Control Units, Dynamometers and
Electronic Speed and Voltage Control Equipment.
Our design staff will be pleased to offer advice
on customers' special requirements.
Synchronous Rotary Converter AC/DC—DC/AC
This E.D.C.C. equipment is installed in the up-to-date laboratory of the Wolverhampton and Staffordshire Technical College
ELECTRO DYNAMIC CONSTRUCTION CO LTD
ST. MARY CRAY · KENT
Tel: ORPINGTON 27551. Grams: "ELEDAMIC, ST. MARY CRAY"
Control Gear Division: Bridgwater, Somerset. Tel: Bridgwater 2882. Glasgow Office: 40 Houldsworth St., C.3. Tel: Central 2620

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. ELECTRONIC AND COMMUNICATION ENGINEERING (INCLUDING RADIO ENGINEERING), JULY 1959

CONTENTS

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.

Example.—Smith, J.: 'Reflections from the Ionosphere', *Proceedings I.E.E.*, Paper No. 4001 R, December, 1954 (102 B, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* COVent Garden 1871. *Telegrams:* 'Voltampere, Phone, London.'
Printed by Unwin Brothers Limited, Woking and London.